**PCT**

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **C12N 15/41, C07K 14/095, A61K 39/125, G01N 33/53, 33/569, C12Q 1/68** — A1

(11) International Publication Number: **WO 97/22701**

(43) International Publication Date: 26 June 1997 (26.06.97)

(21) International Application Number: PCT/AU96/00815

(22) International Filing Date: 18 December 1996 (18.12.96)

(30) Priority Data:
PN 7201 18 December 1995 (18.12.95) AU

(71) Applicant *(for all designated States except US)*: THE UNIVERSITY OF MELBOURNE [AU/AU]; Grattan Street, Parkville, VIC 3052 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: STUDDERT, Michael, J. [AU/AU]; Centre for Equine Virology, School of Veterinary Science, University of Melbourne, Parkville, VIC 3052 (AU). CRABB, Brendan, S. [AU/AU]; Centre for Equine Virology, School of Veterinary Science, University of Melbourne, Parkville, VIC 3052 (AU). FENG, Li [AU/AU]; Centre for Equine Virology, School of Veterinary Science, University of Melbourne, Parkville, VIC 3052 (AU).

(74) Agent: CARTER SMITH & BEADLE; Qantas House, 2 Railway Parade, Camberwell, VIC 3124 (AU).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*

(54) Title: EQUINE RHINOVIRUS 1 PROTEINS

(57) Abstract

Equine rhinovirus 1 (ERhV1) is a respiratory pathogen of horses which has an uncertain taxonomic status. The nucleotide sequence of the ERhV1 genome and amino acid sequence have been substantially determined (figure 2). The predicted polyprotein was encoded by 6,741 nucleotides and possessed a typical picornavirus proteolytic cleavage pattern, including a leader polypeptide. The genomic structure and predicted amino acid sequence of ERhV1 were more similar to those of foot-and-mouth disease viruses (FMDV), the only members of the aphthovirus genus, than other picornaviruses. Nucleotide sequences coding for the complete polyprotein, the polymerase, and VP1 were analyzed separately. The phylogenetic trees confirmed that ERhV1 was more closely related to aphthoviruses than to other picornaviruses. Virion proteins and virus-like particles are described and probes, primers, antigens, vectors, diagnostics and tests developed.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AM | Armenia | GB | United Kingdom | MW | Malawi |
| AT | Austria | GE | Georgia | MX | Mexico |
| AU | Australia | GN | Guinea | NE | Niger |
| BB | Barbados | GR | Greece | NL | Netherlands |
| BE | Belgium | HU | Hungary | NO | Norway |
| BF | Burkina Faso | IE | Ireland | NZ | New Zealand |
| BG | Bulgaria | IT | Italy | PL | Poland |
| BJ | Benin | JP | Japan | PT | Portugal |
| BR | Brazil | KE | Kenya | RO | Romania |
| BY | Belarus | KG | Kyrgystan | RU | Russian Federation |
| CA | Canada | KP | Democratic People's Republic of Korea | SD | Sudan |
| CF | Central African Republic | | | SE | Sweden |
| CG | Congo | KR | Republic of Korea | SG | Singapore |
| CH | Switzerland | KZ | Kazakhstan | SI | Slovenia |
| CI | Côte d'Ivoire | LI | Liechtenstein | SK | Slovakia |
| CM | Cameroon | LK | Sri Lanka | SN | Senegal |
| CN | China | LR | Liberia | SZ | Swaziland |
| CS | Czechoslovakia | LT | Lithuania | TD | Chad |
| CZ | Czech Republic | LU | Luxembourg | TG | Togo |
| DE | Germany | LV | Latvia | TJ | Tajikistan |
| DK | Denmark | MC | Monaco | TT | Trinidad and Tobago |
| EE | Estonia | MD | Republic of Moldova | UA | Ukraine |
| ES | Spain | MG | Madagascar | UG | Uganda |
| FI | Finland | ML | Mali | US | United States of America |
| FR | France | MN | Mongolia | UZ | Uzbekistan |
| GA | Gabon | MR | Mauritania | VN | Viet Nam |

# EQUINE RHINOVIRUS 1 PROTEINS

## INTRODUCTION TO INVENTION

This invention relates to the equine rhinovirus 1 (ERhV1) which has been sequenced and characterized. In particular, the invention relates to nucleotide and protein sequences of ERhV1 and a range of clinical and diagnostic products derived from ERhV1.

## BACKGROUND OF INVENTION

Equine rhinovirus 1 (ERhV1) was first isolated from horses in the United Kingdom and subsequently from horses in mainland Europe, the USA and Australia. Most isolates were from the nasopharynx of horses with an acute, febrile respiratory disease. Virions had the characteristic size and morphology of picornaviruses and were acid-labile. Two other serologically distinct, acid-labile picornaviruses, ERhV2 and ERhV3, have also been isolated from horses.

Considerable uncertainty has surrounded the classification of ERhV1. Physicochemical studies have shown that the nucleic acid density and base composition of ERhV1 differ from those of rhinoviruses. In contrast to rhinoviruses, ERhV1 has a broad host-cell range in vitro and in vivo and there is no evidence of extensive antigenic variation. Infection of horses with ERhV1 causes a disease characterized by an acute febrile respiratory disease accompanied by anemia, fecal and urine shedding and viral persistence. The signs of systemic infection and persistence are not characteristic of rhinovirus infections in other species. The known host range of ERhV1 is broad and includes rabbits, guinea pigs, monkeys and humans, although in these species the virus does not appear to spread horizontally. There is both experimental and epidemiological evidence of ERhV1 infection of humans. A human volunteer inoculated intranasally with ERhV1 developed severe pharyngitis, lymphadenitis, fever and viremia, and high ERhV1 antibody titers were found in the sera of 3 of 12 stable workers whereas no ERhV1 antibody was found in the sera of 159 non-stable workers.

In order to clarify the taxonomic status of ERhV1, a detailed study was undertaken to determine the nucleotide and amino acid sequence of ERhV1. The resultant studies provided the complete nucleotide sequence of the gene encoding

the ERhV1 polyprotein and the 3'-nontranslated region (NTR) as well as part of the nucleotide sequence of the 5'NTR. The amino acid sequence of the various ERhV1 proteins was deduced from the nucleotide sequence.

The analysis of the nucleotide sequence of ERhV1 confirmed previous studies which indicated that many properties of ERhV1 are not consistent with those of other members of the genus *Rhinovirus*. Indeed many of the physicochemical and biological properties of ERhV1 have suggested ERhV1 is more closely related to foot-and-mouth disease virus (FMDV) the sole member of the *Alpthovirus* genus. In addition to the overall sequence similarity, several features of the ERhV1 genome are similar to those of FMDV. The ERhV1 L protein is most similar to its counterpart in aphthoviruses in both length, 207 amino acids in ERhV1 and 201 in FMDV, and in amino acid sequence identity. In aphthoviruses, the L protein catalyses its own cleavage from the polyprotein, and mediates cleavage of the p220 component of the cap-binding complex leading to inhibition of translation of capped mRNAs. Cardiovirus L proteins are only 67-76 amino acids long and are not auto catalytic. In contrast to the cardioviruses, aphthoviruses utilize two distinct initiation codons, which results in different forms of the L protein, Lab and Lb, differing from each other by 28 amino acids at their N-termini.

The second initiation codon occurs in a more favourable context, which is presumably the reason why Lb, the smaller of the two proteins, is the predominant species. Thus far, differences in the function of the two FMDV L proteins have not been detected. ERhV1 also possesses a second ATG, 63 bases downstream from the first optimal ATG, which is also present in a context optimal for initiation of translation. Translation from this ATG would result in an L protein with 21 fewer amino acids at its N-terminus. Therefore, it is probable that ERhV1 possesses a second species of L protein, similar to the FMDV Lb protein. If so, the reason for the existence and conservation of two forms of the L protein in ERhV1 and FMDV is an intriguing question. Curiously, ERhV1 has tand mly repeated ATG codons at each of the possible initiation sites, where the first ATG in each case does not

occur in a context optimal for translation. The role of these ATGs may be to ensure that translation is initiated from both possible initiation sites.

The 2A protease is only 16 amino acids in length in both FMDV and ERhV1, compared to 142-149 amino acids in other picornaviruses. In FMDV 2A protease cleaves at its C-terminus but, unlike the 2A protease of other picornaviruses, appears not to have a role in shut down of host cell macromolecular synthesis. The high degree of conservation of the FMDV and ERhV1 2A proteins is intriguing and suggests an important role for this protein in the diseases produced by these viruses.

It may be expected that the tree derived from the complete polyprot in coding sequence would provide the most representative view of the taxonomic status of ERhV1 by reducing any bias imparted by using restricted parts of the genome with highly variable evolutionary rates. However, such analysis is restricted because there are only a few complete polyprotein sequences available. The polymerase genes are the most conserved genes in positive strand RNA viruses and they have been used to construct a taxonomy, and to predict the ancient roots, of these viruses. In contrast to the polymerase gene, the VP1 gene encodes the major antigenic determinants of the virus and evolves more rapidly than other regions in the genome. The diversity of VP1 regions make them useful for the study of closely related picornaviruses. Thus, trees based on the polymerase and VP1 genes presumably reflect the extremes of evolutionary rates from which the taxonomic status and evolutionary origin of ERhV1 could be identified. The ERhV1 VP1 amino acid sequence was more similar to FMDV than to any other sequence in the data base; this was true even when representative segments across the entire sequence were separately analysed.

Therefore, we consider that the difference in the topology of the VP1, compared to the other two trees, is most unlikely to be a consequence of gen tic recombination. The topographic differences between the three ERhV1 trees compared to those of aphthoviruses, particularly the VP1 derived trees, as well as the presence of only one VPg gene in ERhV1 genome. leads us to conclude that

ERhV1 is probably a member of a distinct genus proposed to be called *Equirhinovirus*.

The reassessment of the taxonomic status of ERhV1 focuses on a requirement to reassess the biology of the virus particularly with respect to the nature of clinical disease as well as means for control by vaccination and improved methods of diagnosis. For example, cardioviruses and aphthoviruses cause viremic infections accompanied by myocarditis. Clinical disease caused by ERhV1 is generally considered to be confined to the respiratory tract even though there is a viremia and the virus is shed in faeces and urine. Whether ERhV1 infection produces systemic disease similar to that observed in aphthovirus or cardiovirus infections, including the production of myocarditis, needs to be investigated. There is serological evidence that the incidence of ERhV1 infection is as high as 50% in some horse populations however, the number of reported isolations of ERhV1 is very small. We have clear evidence that primary isolation of the virus from clinical specimens is known to be difficult, suggesting that the true incidence of ERhV1 disease is much greater than reported.

The determination of the complete nucleotide sequence of ERhV1 polyprotein has important practical applications in developing novel methods for the diagnosis and control of ERhV disease in horses and other species.

## OBJECT AND STATEMENT OF INVENTION

In one aspect, the invention provides a substantially pure nucleotide sequence for ERhV1 being:

a substantially pure nucleotide sequence for ERhV1 being:

```
CCGTCAAGCC CGTTGCCTGT ATAGCCAGGT AACCGGACAG CGGCTTGCTG GATTTTCCCG   -375
GTGCCATTGC TCTGGATGGT GTCACCAAGC TGACAAATGC GGAGTGAACC TCACAAAGCG   -315
ACACGCCTGT GGTAGCGCTG CCCAAAAGGG AGCGGAACTC CCCGCCGAGG CGGTCCTCTC   -255
TGGCCAAAAG CCCAGCGTTG ATAGCGCCTT TTGGGATGCA GGAACCCCAC CTGCCAGGTG   -195
TGAAGTGGAG TGAGCGGATC TCCAATTTGG TCTGTTCTGA ACTACACCAT TTACTGCTGT   -135
GAAGAATGCC CTGGAGGCAA GCTGGTTACA GCCCTGACCA GGCCCTGCCC GTGACTCTCG    -75
ACCGGCGCAG GGTCAAAAAT TGTCTAAGCA GCAGCAGGAA CGCGGGAGCG TTTCTTTTCC    -15
TTTTGTACTG ACATGATGGC GGCGTCTAAG GTGTATAGAG TTTGCGAGCA GACTCTGCTG     45
GCAGGTGCCG TTCGCATGAT GGACAAATTC TTGCAAAAGA GAACTGTTTT TGTCCCCCAT    105
CTTGACAAAA CAATTCGTTT GACTGGACTC CACAATTATG ACAATACTTG CTGGTTGAAT    165
GCCTTGACAC AACTGACACA GATTCTTGGA ATTCGGCTTT TTGATGAACA CTTCGGCAAT    225
AGAGGTCTGT TCACTCGGAA AACAATTGAT TGGGTGAGTG ACCAGACTGG TATAAAAGAT    285
```

CTAAAATCAG GAGCACCGCC ACTCGTGGTG GTGTACAAAC TGTGGCAACA TGGACACTTG 345
GATGTCGGTA CGATGGAGAA ACCCCGGTCG ATTACTCTAT GGTCTGGCCC CAAAGTGTGT 405
CTTTCTGATT TCTGGGCCTG TGTTTCGGCA AAACCGGGAC ATGCAGTATT CTACCTTCTC 465
ACAAGCGAGG GTTGGATCTG TGTTGATGAC AAGAAAATAT ACCCAGAAAC ACCCAAAACA 525
GAGGATGTAC TTGTTTTTGC GCCCTATGAC TTTGAGTCAC TGGGCAAGGA CCCACCAAAG 585
CTACACCAGA GATATGAAAA AGCATTTGAG CTCAGTGGCG GAGGTACATC CACTCCAACA 645
ACTGGCAACC AAAACATGTC CGGAAACAGT GGTTCAATTG TTCAAAATTT TTACATGCAA 705
CAGTACCAGA ATTCAATTGA CGCAGACCTG GGAGACAATG TGATTAGCCC TGAAGGCCAG 765
GGCAGCAACA CTAGTAGTTC AACCTCATCA AGCCAATCCT CTGGCTTGGG CGGGTGGTTC 825
TCTAGTTTGC TGAACCTTGG AACAAAACTA CTGGCTGACA AGAAGACAGA AGAGACTACA 885
AACATTGAAG ACAGAATTGA AACAACAGTG GTTGGAGTCA CTATTATTAA TTCACAAGGA 945
TCTGTTGGAA CAACCTACTG TTACTCCAAA CCGGATGGTA GACCACCATC CACAGTGTCA 1005
GACCCAGTTA CCAGACTTGG ACCCACGCTT TCCAGGCACT ACACATTTAA GGTAGGTGAG 1065
TGGCCCCATT CTCAATCACA TGGTCACGCA TGGATCTGTC CGTTGCCAGG TGACAAACTC 1125
AAGAAGATGG GCAGTTTTCA TGAGGTTGTC AAAGCCCACC ACCTGGTCAA GAACGGCTGG 1185
GATGTGGTTG TGCAGGTGAA TCCCTCATTT GCTCACTCCG GGCCGCTGTG TGTAGCAGCA 1245
GTGCCGGAGT ACGAACACAC ACATGAGAAA GCACTCAAGT GGTCTGAGCT TGAGGAACCA 1305
GCTTACACAT ACCAACAACT TTCAGTTTTT CCCCACCAGT TGCTAAATTT GAGGACAAAT 1365
TCATCAGTGC ATTTGGTGAT GCCCTACATT GGGCCAGGCC AACCAACAAA TCTGACTTTG 1425
CACAACCCGT GGACCATTGT TATTTTAATT TTGTCTGAAT TGACAGGACC TGGCCAAACT 1485
GTGCCTGTGA CCATGTCGGT GGCTCCCATC GATGCAATGG TTAATGGGCC TCTTCCAAAT 1545
CCAGAGGCAC CGATTAGAGT GGTGTCTGTG CCTGAATCAG ATTCTTTTAT GTCTTCAGTA 1605
CCTGATAATT CGACTCCACT ATACCCCAAG GTTGTGGTCC CACCGCGCCA AGTTCCTGGC 1665
CGGTTTACAA ATTTCATTGA TGTGGCAAAA CAGACATATT CATTTTGTTC CATTTCTGGA 1725
AAACCTTATT TTGAGGTTAC CAACACCTCT GGGGACGAGC CACTGTTTCA GATGGATGTG 1785
TCGCTCAGTG CGGCAGAGCT ACATGGCACT TACGTAGCTA GTTTGTCATC ATTTTTTGCA 1845
CAGTACAGAG GCTCACTTAA TTTCAACTTT ATTTTCACTG GTGCAGCAGC CACTAAGGCA 1905
AAGTTTCTGG TTGCTTTTGT GCCTCCCCAC AGTGCAGCGC CCAAAACGCG CGATGAAGCA 1965
ATGGCGTGCA TCCATGCCGT GTGGGATGTT GGCTTGAACT CAGCTTTTTC TTTTAATGTA 2025
CCTTATCCCT CCCCTGCTGA CTTCATGGCC GTTTATTCTG CGGAACGGAC GGTTGTGAAT 2085
GTCTCTGGAT GGCTTCAAGT TTATGCACTA ACAGCTCTAA CTTCAACTGA CATTGCCGTG 2145
AACAGTAAAG GCCGTGTGCT GGTTGCTGTT TCCGCCGGCC CAGACTTCTC CCTTCGTCAC 2205
CCGGCGGACC TGCCCGACAA GCAGGTTACC AATGTGGGAG AGGATGGTGA ACCCGGTGAG 2265
ACAGAGCCTC GTCATGCTTT GTCACCCGTG GACATGCACG TGCACACAGA TGTCAGTTTC 2325
TTGCTTGACC GGTTCTTTGA TGTTGAGACA CTTGAGCTTT CAAATTTGAC AGGTTCTCCT 2385
GCCACACATG TTCTGGATCC GTTTGGCTCG ACTGCCCAAC TGGCTTGGGC ACGTCTGCTA 2445
AACACTTGCA CCTACTTCTT TTCTGATTTG GAATTGTCAA TCCAGTTTAA ATTTACCACC 2505
ACTCCGTCCT CTGTTGGAGA GGGCTTTGTG TGGGTGAAGT GGCTCCCTGT TGGAGCACCA 2565
ACCAAGACCA CAGATGCTTG GCAGTTAGAA GGAGGTGGAA ATTCAGTTAG AATTCAAAAA 2625
TTGGCCGTTG CAGGGATGTG CCCCACTGTT GTGTTCAAGA TTGCAGGCTC CCGTTCACAA 2685
GCCTGTGCTT CAGCGTTGCC ATATACATCA ATGTGGCGTG TTGTGCCAGT CTTTTACAAT 2745
GGCTGGGGTG CACCTACCAA AGAAAAGGCA ACCTACAATT GGCTTCCTGG TGCACACTTT 2805
GGTTCCATCT TGCTGACTTC TGATGCGCAT GATAAAGGAG GGTGCTACTT GCGGTATGCT 2865
TTCCGCGCGC CAGCGATGTA TTGCCCTCGA CCCATTCCGC CGGCTTTTAC GCGTCCAGCG 2925
GACAAAACCA GACATAAATT TCCCACTAAC ATCAACAAAC AGTGTACTAA TTACTCTCTC 2985
CTCAAATTGG CTGGAGATGT TGAGAGCAAC CCTGGCCCCA CTATTTTTTC CAAAGCATCA 3045
GCAGACCTGA ATGCCTTGTC AACGTCGCTA GGTGAATTGA CTGGCATGCT AAAAGATCTT 3105

AAAGCCAAGG CAGAAACTTA TTCCCCGTTT TACAAAATGG CCAAAATGCT TTTCAAACTT 3165
GCAACACTAG CTGTGGCAGC TATGAGGACA AAGGACCCAG TAGTGGTGGT TATGTTGATT 3225
GCTGATTTCG GATTGGAGGT CTTTGACACT GGGTTTTTCT TTTCCTACTT TCAAGAGAAG 3285
TTGCAGCCTT ATATGAAAAC TATTCCTGGT AAGATTTCTG ATTTGGTCAC TGATGCGGCT 3345
ACGGCTGCCG CCCAAATTCC AAAGGGAGTG TATTCTTTTG TGTCGTCATT TTTCGAAACG 3405
CCTGAAGGAG TGGTTGAGAA GCAGGTGTCT CTTCGGACAG TGAATGACAT ATTTGCTTTG 3465
CTTAAAAATT CTGATTGGTT CATAAAGACT CTTGTTGCCC TCAAGAAATG GCTGACATCC 3525
TGGTTTGCTC AAGAACAACA GGCAGATGAT GCGCTCTATT CAGAATTGGA AAAATATCCC 3585
TTGTACAAGT TAAAATTGAA GGAACCTGAT ACTCAAGAGG AAGCGCGCCA GTGGTTTAAA 3645
GACATGCAGC AGCGTGCTCT CGCTGTGAAG GACAAAGGTC TCTTTTCCCT CCTGCAAATT 3705
CCATTAGTTA ACTTGCCCCA GAGCCGTCCA GAGCCCGTTG TATGCGTCCT TCGGGGCGCA 3765
TCAGGGCAAG GCAAATCTTA TTTGGCAAAT CTGATGGCTC AAGCAATTTC GCTTCTCTTG 3825
GTTGGCAAGC AGGACAGTGT GTGGAGTTGT CCTCCTGACC CCACATATTT TGATGGCTAT 3885
AACGGACAGG CTGTGGTGAT TATGGATGCA TTGGGCCAGG ATCCGAATGG TGCTGACTTT 3945
AAATATTTTT GCCAGATGGT CTCTACAACA GCTTTTGTAC CACCTATGGC CCATTTGGAT 4005
GATAAAGGCA TTCCATTTAC TTCTCCTGTT GTTATTTGTA CTACAAATTT GCATTCATCT 4065
TTTACCCCTA TTACTGTTTC TTGTCCTGAA GCTCTTAAGA GGAGGTTTCG GTTTGATGTG 4125
ACGGTGTCCG CTAAACCGGG CTTTGTGCGC ACTGTTGGTT CAAACCAGCT TTTGAATCTC 4185
CCACTTGCTC TTAAGCCAGC TGGTCTTCCC CCACACCCTA TCTTTGAAAA TGACATGCCC 4245
ATTATAAATG GGCAGGCTGT TAAATTGGCT CTTTCTGGTG GAGAAGTGAC AGCTTTTGAG 4305
CTTATTGAGA TGATACTGTC AGAAGTTCAA AACAGACAAG ACACACACAA AATGCCCATT 4365
TTTAAACAAT CATGGTCTGA TTTGTTCAGA AAGTGTACAA CTGATGAGGA ACAGAAAATG 4425
TTGCAGTTTT TAATTGACAA TAAAGATTCA GAAATTCTCA GGGCGTTTGT TTCAGAACGC 4485
TCCATTTTAC TACATGAAGA GTATCTTAAA TGGGAGTCAT ATATGACCAG GAGAGCCAAG 4545
TTTCACCGCC TGGCTGCTGA TTTTGCTATG TTTCTATCCA TTCTTACTTC ACTGATTGTT 4605
ATTTTTTGTT TAGTTTATTC TATGTATCAA CTTTTTAAGA CCCCTGACGA GCAATCAGCT 4665
TATGATCCTT CAACTAAGCC AAAACCAAAG ACCCAGGAAG TGAAAACACT GAAGATTAGG 4725
ACTGAGACTG GTGTACCAGC AACTGACTTG CAACAATCCA TCATGAAAAA TGTTCAGCCA 4785
ATTGAGCTTT ACCTTGACAA TGAATTGGTT ACTGACTGCT CTGCCTTGGG TGTTTATGAC 4845
AATTCATATT TGGTGCCCCT TCATTTGTTT GAATTTGATT TTGATACCAT TGTGCTTGGT 4905
GGACGTCATT ACAAGAAAGC TGAGTGTGAG AAGGTAGAGT TTGAGCTTGA AGTGAATGGA 4965
GACGTGGTGT CATCAGATGC GTGTCTACTT CGAGTGTCAT CGGGGCCTAA AGTTAGAAAT 5025
ATTGTTCATC TTTTTACAAA TGAAATTGAA TTGAAGAAAA TGACCCAAGT GACAGGAATC 5085
ATGAATTCAC CACACCAGGC ACGCACTGTG TTTTTTGGCA GTTTTTTGAC AGTGAGGAAG 5145
TCCATCTTAA CATCGGATGG GACTGTAATG CCCAATGTTT TGTCCTATGC CGCTCAGACC 5205
TCGCGTGGGT ATTGTGGCGC TGCAATTGTT GCTGGCTCAC CTGCCCGCAT AATTGGTATC 5265
CATTCAGCTG GCACTGGATC TGTTGCATTT TGCTCCCTGG TGTCCAGAGA CGCGCTGGAG 5325
CAACTCTGGC CCCAGAAACA GGGCAACGTT AGTCGCCTTG ATGACGATGT GAGGGTGTCT 5385
GTTCCGCGCC GCTCCAAATT GGTGAAATCA TTGGCTTACC CCATTTTCAA ACCTGACTAT 5445
GGCCCAGCGC CACTCTCTCA ATTTGACAAG CGCCTGTCAG ACGGCGTGAA GCTGGATGAA 5505
GTGGTTTTTG CTAAACATAC TGGAGACAAG GAGATTTCCG CACAGGACCA GAAATGGCTC 5565
TTGCGTGCGG CGCATGTATA CGCCCAGAAG GTTTTCTCCC GGATTGGATT TGACAACCAG 5625
GCTTTGACTG AAAAAGAGGC CATTTGTGGC ATTCCTGGCC TTGACAAGAT GGAGCAGGAC 5685
ACCGCTCCCG GGCTGCCCTA TGCTCAGCAA AATAAGAGAA GGAAAGACAT CTGTGATTTT 5745
GAAGAGGGCC GGCTGAAGGG CGCCGAACTC CAAAAGGACA GATTTATGGC TGGTGACTAC 5805
TCTAATTTGG TCTATCAATC ATTTTTGAAA GATGAGATCC GCCCACTTGA GAAAGTTAGG 5865
GCTGGAAAGA CCCGCCTGAT TGACGTGCCG CCGATGCCCC ATGTGGTGGT TGGTAGGCAG 5925

CTCTTGGGCC GGTTTGTGGC AAAATTTCAT GAAGCAAATG GATTTGACAT TGGCTCAGCC 5985
ATTGGATGTG ACCCAGATGT GGACTGGACT CGGTTTGGCC TCGAGTTGGA GCGTTTCAGG 6045
TATGTATATG CCTGTGACTA CTCACGGTTC GATGCCAACC ATGCAGCTGA TGCAATGAGA 6105
GTTGTGCTTA ACTACTTTTT CTCTGAGGAC CACGGTTTCG ACCCTGGTGT GCCTGCTTTT 6165
ATTGAGTCAC TGGTTGATTC AGTGCATGCC TATGAAGAGA AAAGGTATAA CATCTACGGT 6225
GGCTTGCCAT CCGGGTGTTC CTGCACATCA ATTTTGAATA CCATCTTGAA CAATGTTTAC 6285
ATTCTTGCAG CTATGATGAA GGCTTATGAG AATTTTGAGC CAGATGACAT TCAGGTCATT 6345
TGCTATGGGG ACGACTGCCT CATTGCTTCT GATTTTGAAA TTGATTTCCA ACAACTGGTG 6405
CCTGTCTTTT CTAGTTTTGG ACAGGTAATA ACTACAGCTG ACAAGACTGA TTTTTTTAAA 6465
CTGACAACGC TTTCGGAGGT GACCTTCCTT AAGCGCGCTT TTGTTCTGAC GGCCTTTTAC 6525
AAGCCAGTGA TGGATGTGAA GACCCTTGAA GCAATCTTAA GCTTTGTTCG CCCAGGCACA 6585
CAGGCTGAAA AGCTCCTGTC CGTGGCGCAG TTGGCAGGCC ACTGCGAACC GGAGCAGTAT 6645
GAGCGCCTGT TTGAGCCCTT TGCTGGGATG TATTTCGTCC CTACTTGGCG ACTTGCGCCT 6705
GCAGTGGTTG ATGAAGCTTG GATGCTAAAT TCTTTTTGAC TTTGTTTTTC TTTGTTTTCT 6765
TTTAGGCTTT TAAGGTGTTA AGTTTAAAGG TTAAGAGTTT TTAGAAGTTA AGATAGAGTT 6825
TAGTTTTTAG TTTTGAGC-poly(A)

as disclosed in Fig. 2 and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants, degeneracy equivalents and deletion mutants thereof.

In another aspect, the invention provides a substantially pure amino acid sequence being:

a substantially pure amino acid sequence being:

M A A S K V Y R V C E Q T L L A G A V R M M D K F
L Q K R T V F V P H L D K T I R L T G L H N Y D N
T C W L N A L T Q L T Q I L G I R L F D E H F G N
R G L P T R K T I D W V S D Q T G I K D L K S G A
P P L V V V Y K L W Q H G H L D V G T M E K P R S
I T L W S G P K V C L S D F W A C V S A K P G H A
V F Y L L T S E G W I C V D D K K I Y P E T P K T
E D V L V F A P Y D F E S L G K D P P K L H Q R Y
L ↓ VP4
E K A F E L S G G G T S T P T T G N Q N M S G N S
G S I V Q N F Y M Q Q Y Q N S I D A D L G D N V I
S P E G Q G S N T S S S T S S S Q S S G L G G W F
VP4 ↓ VP2
S S L L N L G T K L L A D K K T E E T T N I E D R
I E T T V V G V T I I N S Q G S V G T T Y C Y S K
P D G R P P S T V S D P V T R L G P T L S R H Y T

F K V G E W P H S Q S H G H A W I C P L P G D K L
K K M G S F H E V V K A H H L V K N G W D V V V Q
V N P S F A H S G P L C V A A V P E Y E H T H E K
A L K W S E L E E P A Y T Y Q Q L S V F P H Q L L
N L R T N S S V H L V M P Y I G P G Q P T N L T L
H N P W T I V I L I L S E L T G P G Q T V P V T M

VP2 ↓ VP3

S V A P I D A M V N G P L P N P E A P I R V V S V
P E S D S F M S S V P D N S T P L Y P K V V V P P
R Q V P G R F T N F I D V A K Q T Y S F C S I S G
K P Y F E V T N T S G D E P L F Q M D V S L S A A
E L H G T Y V A S L S S F F A Q Y R G S L N F N F
I F T G A A A T K A K F L V A F V P P H S A A P K
T R D E A M A C I H A V W D V G L N S A F S F N V
P Y P S P A D F M A V Y S A E R T V V N V S G W L
Q V Y A L T A L T S T D I A V N S K G R V L V A V

VP3 ↓ VP1

S A G P D F S L R H P A D L P D K Q V T N V G E D
G E P G E T E P R H A L S P V D M H V H T D V S F
L L D R F F D V E T L E L S N L T G S P A T H V L
D P F G S T A Q L A W A R L L N T C T Y F F S D L
E L S I Q F K F T T T P S S V G E G F V W V K W L
P V G A P T K T T D A W Q L E G G G N S V R I Q K
L A V A G M C P T V V F K I A G S R S Q A C A S A
L P Y T S M W R V V P V F Y N G W G A P T K E K A
T Y N W L P G A H F G S I L L T S D A H D K G G C
Y L R Y A F R A P A M Y C P R P I P P A F T R P A

VP1 ↓ 2A

D K T R H K F P T N I N K Q C T N Y S L L K L A G

2A ↓ 2B

D V E S N P G P T I F S K A S A D L N A L S T S L
G E L T G M L K D L K A K A E T Y S P F Y K M A K
M L F K L A T L A V A A M R T K D P V V V V M L I
A D F G L E V F D T G F F F S Y F Q E K L Q P Y M
K T I P G K I S D L V T D A A T A A A Q I P K G V

2B ↓ 2C

Y S F V S S F F E T P E G V V E K Q V S L R T V N
D I F A L L K N S D W F I K T L V A L K K W L T S
W F A Q E Q Q A D D A L Y S E L E K Y P L Y K L K
L K E P D T Q E E A R Q W F K D M Q Q R A L A V K
D K G L F S L L Q I P L V N L P Q S R P E P V V C
V L R G A S G Q G K S Y L A N L M A Q A I S L L L
V G K Q D S V W S C P P D P T Y F D G Y N G Q A V
V I M D A L G Q D P N G A D F K Y F C Q M V S T T
A F V P P M A H L D D K G I P F T S P V V I C T T
N L H S S F T P I T V S C P E A L K R R F R F D V
T V S A K P G F V R T V G S N Q L L N L P L A L K
P A G L P P H P I F E N D M P I I N G Q A V K L A
L S G G E V T A F E L I E M I L S E V Q N R Q D T

2C ↓ 3A

H K M P I F K Q S W S D L F R K C T T D E E Q K M
L Q F L I D N K D S E I L R A F V S E R S I L L H
E E Y L K W E S Y M T R R A K F H R L A A D F A M
F L S I L T S L I V I F C L V Y S M Y Q L F K T P

3A ↓ 3B

D E Q S A Y D P S T K P K P K T Q E V K T L K I R

3B ↓ 3C

T E T G V P A T D L Q Q S I M K N V Q P I E L Y L
D N E L V T D C S A L G V Y D N S Y L V P L H L F
E F D F D T I V L G G R H Y K K A E C E K V E F E
L E V N G D V V S S D A C L L R V S S G P K V R N
I V H L F T N E I E L K K M T Q V T G I M N S P H
Q A R T V F F G S F L T V R K S I L T S D G T V M
P N V L S Y A A Q T S R G Y C G A A I V A G S P A
R I I G I H S A G T G S V A F C S L V S R D A L E

3C ↓ 3D

Q L W P Q K Q G N V S R L D D D V R V S V P R R S
K L V K S L A Y P I F K P D Y G P A P L S Q F D K
R L S D G V K L D E V V F A K H T G D K E I S A Q
D Q K W L L R A A H V Y A Q K V F S R I G F D N Q

A L T E K E A I C G I P G L D K M E Q D T A P G L
P Y A Q Q N K R R K D I C D F E E G R L K G A E L
Q K D R F M A G D Y S N L V Y Q S F L K D E I R P
L E K V R A G K T R L I D V P P M P H V V V G R Q
L L G R F V A K F H E A N G F D I G S A I G C D P
D V D W T R F G L E L E R F R Y V Y A C D Y S R F
D A N H A A D A M R V V L N Y F F S E D H G F D P
G V P A F I E S L V D S V H A Y E E K R Y N I Y G
G L P S G C S C T S I L N T I L N N V Y I L A A M
M K A Y E N F E P D D I Q V I C Y G D D C L I A S
D F E I D F Q Q L V P V F S S F G Q V I T T A D K
T D F F K L T T L S E V T F L K R A F V L T A F Y
K P V M D V K T L E A I L S F V R P G T Q A E K L
L S V A Q L A G H C E P E Q Y E R L F E P F A G M

3D

Y F V P T W R L A P A V V D E A W M L N S F

as disclosed in Fig. 2.

In another aspect, the invention provides proteins derived from ERhV1 which exhibit virus like particle characteristics incorporating VP1 and having the following amino acid sequence:

a protein or virus like particle incorporating VP1, derived from ERhV1 and having the following amino acid sequence:

V T N V G E D G E P G E T E P R H A L S P V D M H
V H T D V S F L L D R F F D V E T L E L S N L T G
S P A T H V L D P F G S T A Q L A W A R L L N T C
T Y F F S D L E L S I Q F K F T T T P S S V G E G
F V W V K W L P V G A P T K T T D A W Q L E G G G
N S V R I Q K L A V A G M C P T V V F K I A G S R
S Q A C A S A L P Y T S M W R V V P V F Y N G W G
A P T K E K A T Y N W L P G A H F G S I L L T S D
A H D K G G C Y L R Y A F R A P A M Y C P R P I P
P A F T R P A D K T R H K F P T N I N K Q C T

In another aspect. the invention provides proteins derived from ERhV1 which exhibit virus like particle characteristics incorporating VP2 and having the following amino acid sequence:

a protein or virus like particle incorporating VP2. derived from ERhV1 and having the following amino acid sequence:

D K K T E E T T N I E D R I E T T V V G V T I I N
S Q G S V G T T Y C Y S K P D G R P P S T V S D P
V T R L G P T L S R H Y T F K V G E W P H S Q S H
G H A W I C P L P G D K L K K M G S F H E V V K A
H H L V K N G W D V V V Q V N P S F A H S G P L C
V A A V P E Y E H T H E K A L K W S E L E E P A Y
T Y Q Q L S V F P H Q L L N L R T N S S V H L V M
P Y I G P G Q P T N L T L H N P W T I V I L I L S
E L T G P G Q T V P V T M S V A P I D A M V N G P
L P N P E

In another aspect, the invention provides proteins derived from ERhV1 which exhibit virus like particle characteristics incorporating VP3 and having the following amino acid sequence:

a protein or virus like particle incorporating VP3, derived from ERhV1 and having the following amino acid sequence:

A P I R V V S V P E S D S F M S S V P D N S T P L
Y P K V V V P P R Q V P G R F T N F I D V A K Q T
Y S F C S I S G K P Y F E V T N T S G D E P L F Q
M D V S L S A A E L H G T Y V A S L S S F F A Q Y
R G S L N F N F I F T G A A A T K A K F L V A F V
P P H S A A P K T R D E A M A C I H A V W D V G L
N S A F S F N V P Y P S P A D F M A V Y S A E R T
V V N V S G W L Q V Y A L T A L T S T D I A V N S
K G R V L V A V S A G P D F S L R H P A D L P D K
Q

In another aspect, the invention provides proteins derived from ERhV1 which exhibit virus like particle characteristics incorporating VP4 and having the following amino acid sequence:

a protein or virus like particle incorporating VP4, derived from ERhV1 and having the following amino acid sequence:

G G G T S T P T T G N Q N M S G N S G S I V Q N F
Y M Q Q Y Q N S I D A D L G D N V I S P E G Q G S
N T S S S T S S S Q S S G L G G W F S S L L N L G
T K L L A

The invention also provides a virus like particle comprising any one or a combination of VP1, VP2, VP3 and VP4.

In another aspect, the invention provides a substantially pure nucleotide sequence for VP1 being:

GTTACCAATG TGGGAGAGGA TGGTGAACCC GGTGAGACAG AGCCTCGTCA TGCTTTGTCA
CCCGTGGACA TGCACGTGCA CACAGATGTC AGTTTCTTGC TTGACCGGTT CTTTGATGTT
GAGACACTTG AGCTTTCAAA TTTGACAGGT TCTCCTGCCA CACATGTTCT GGATCCGTTT
GGCTCGACTG CCCAACTGGC TTGGGCACGT CTGCTAAACA CTTGCACCTA CTTCTTTTCT
GATTTGGAAT TGTCAATCCA GTTTAAATTT ACCACCACTC CGTCCTCTGT TGGAGAGGGC
TTTGTGTGGG TGAAGTGGCT CCCTGTTGGA GCACCAACCA AGACCACAGA TGCTTGGCAG
TTAGAAGGAG GTGGAAATTC AGTTAGAATT CAAAAATTGG CCGTTGCAGG GATGTGCCCC
ACTGTTGTGT TCAAGATTGC AGGCTCCCGT TCACAAGCCT GTGCTTCAGC GTTGCCATAT
ACATCAATGT GGCGTGTTGT GCCAGTCTTT TACAATGGCT GGGGTGCACC TACCAAAGAA
AAGGCAACCT ACAATTGGCT TCCTGGTGCA CACTTTGGTT CCATCTTGCT GACTTCTGAT
GCGCATGATA AAGGAGGGTG CTACTTGCGG TATGCTTTCC GCGCGCCAGC GATGTATTGC
CCTCGACCCA TTCCGCCGGC TTTTACGCGT CCAGCGGACA AAACCAGACA TAAATTTCCC
ACTAACATCA ACAAACAGTG TACT

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

In another aspect, the invention provides a substantially pure nucleotide sequence for VP2 being:

GACAAGAAGA CAGAAGAGAC TACAAACATT GAAGACAGAA TTGAAACAAC AGTGGTTGGA
GTCACTATTA TTAATTCACA AGGATCTGTT GGAACAACCT ACTGTTACTC CAAACCGGAT
GGTAGACCAC CATCCACAGT GTCAGACCCA GTTACCAGAC TTGGACCCAC GCTTTCCAGG
CACTACACAT TTAAGGTAGG TGAGTGGCCC CATTCTCAAT CACATGGTCA CGCATGGATC
TGTCCGTTGC CAGGTGACAA ACTCAAGAAG ATGGGCAGTT TTCATGAGGT TGTCAAAGCC
CACCACCTGG TCAAGAACGG CTGGGATGTG GTTGTGCAGG TGAATCCCTC ATTTGCTCAC
TCCGGGCCGC TGTGTGTAGC AGCAGTGCCG GAGTACGAAC ACACACATGA GAAAGCACTC
AAGTGGTCTG AGCTTGAGGA ACCAGCTTAC ACATACCAAC AACTTTCAGT TTTTCCCCAC
CAGTTGCTAA ATTTGAGGAC AAATTCATCA GTGCATTTGG TGATGCCCTA CATTGGGCCA
GGCCAACCAA CAAATCTGAC TTTGCACAAC CCGTGGACCA TTGTTATTTT AATTTTGTCT
GAATTGACAG GACCTGGCCA AACTGTGCCT GTGACCATGT CGGTGGCTCC CATCGATGCA
ATGGTTAATG GGCCTCTTCC AAATCCAGAG

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

In another aspect, the invention provides a substantially pure nucleotide sequence for VP3 being:

GCACCGATTA GAGTGGTGTC TGTGCCTGAA TCAGATTCTT TTATGTCTTC AGTACCTGAT
AATTCGACTC CACTATACCC CAAGGTTGTG GTCCCACCGC GCCAAGTTCC TGGCCGGTTT
ACAAATTTCA TTGATGTGGC AAAACAGACA TATTCATTTT GTTCCATTTC TGGAAAACCT
TATTTTGAGG TTACCAACAC CTCTGGGGAC GAGCCACTGT TTCAGATGGA TGTGTCGCTC
AGTGCGGCAG AGCTACATGG CACTTACGTA GCTAGTTTGT CATCATTTTT TGCACAGTAC
AGAGGCTCAC TTAATTTCAA CTTTATTTTC ACTGGTGCAG CAGCCACTAA GGCAAAGTTT
CTGGTTGCTT TTGTGCCTCC CCACAGTGCA GCGCCCAAAA CGCGCGATGA AGCAATGGCG
TGCATCCATG CCGTGTGGGA TGTTGGCTTG AACTCAGCTT TTTCTTTTAA TGTACCTTAT
CCCTCCCCTG CTGACTTCAT GGCCGTTTAT TCTGCGGAAC GGACGGTTGT GAATGTCTCT
GGATGGCTTC AAGTTTATGC ACTAACAGCT CTAACTTCAA CTGACATTGC CGTGAACAGT
AAAGGCCGTG TGCTGGTTGC TGTTTCCGCC GGCCCAGACT TCTCCCTTCG TCACCCGGCG
GACCTGCCCG ACAAGCAG

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

In another aspect, the invention provides a substantially pure nucleotide sequence for VP4 being:

GGCGGAGGTA CATCCACTCC AACAACTGGC AACCAAAACA TGTCCGGAAA CAGTGGTTCA
ATTGTTCAAA ATTTTTACAT GCAACAGTAC CAGAATTCAA TTGACGCAGA CCTGGGAGAC
AATGTGATTA GCCCTGAAGG CCAGGGCAGC AACACTAGTA GTTCAACCTC ATCAAGCCAA
TCCTCTGGCT TGGGCGGGTG GTTCTCTAGT TTGCTGAACC TTGGAACAAA ACTACTGGCT

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

In another aspect, the invention provides oligonucleotide primers derived from the nucleotide sequence of Fig. 2 being highly specific for ERhV1 or cross reactive with other ERhV types.

The oligonucleotide primers may have any one of the following nucleotide sequences:

| | | | |
|---|---|---|---|
| VP1F | 5' | GTTGTGTTCAAGATTGCAGGC | 3' |
| VP1R1 | 5' | TTGCTCTCAACATCTCCAGC | 3' |
| VP1R2 | 5' | TAGCACCCTCCTTTATCATGCG | 3' |

In another aspect, the invention provides an oligonucleotide probe derived from the sequence of Fig. 2.

In another aspect, the invention provides diagnostic reagents, methods and kits characterised by the aforesaid oligonucleotide primers and probes.

In another aspect, the invention provides antigens comprising any one or a combination of the non-capsid proteins, being other than the individual VP1 to VP4 proteins, that are cleavage products of the polypeptide of Figure 2.

In another aspect, the invention provides vaccines and vectors incorporating any one or a combination of virion proteins VP1 to VP4.

In another aspect, the invention provides diagnostic tests for the detection of antibodies to ERhV1 in blood of horses or other animals characterised by the use of the aforesaid antigens. Such diagnostic tests may be ELISA based.

In a particularly preferred embodiment, the invention provides a test to distinguish horses infected with ERhV1 in which said virus had replicated from horses which have been vaccinated with the vaccine incorporating any one or a

combination of virion proteins VP1 to VP4; comprising the steps of applying an antigen being any one or a combination of non-capsid proteins, being other than VP1 to VP4, that are cleavage products of the polypeptide of Figure 2 to a horse and testing for an immunoreaction thereto, wherein a positive immunoreaction would indicate that said horse had been infected with ERhV1 and a negative immunoreaction would indicate that said horse has not been infected with ERhV1.

In another aspect, the invention provides recombinant plasmids incorporating nucleotide sequences and subsequences derived from the nucleotide sequences of Fig. 2. The recombinant plasmid may comprise the P1-2A-3C region of the ERhV1 genome.

In another aspect, the invention provides a host system characterised by incorporating the nucleotide sequence of Fig. 2 or part thereof. The host may be E.coli, vaccinia virus, baculovirus or yeast.

In another aspect, the invention provides a process for producing a protein product derived from ERhV1 comprising the steps of selecting out a gene of interest from the ERhV1 nucleotide sequence of Fig. 2 and expressing said protein product in a suitable host system.

## DETAILED DESCRIPTION OF INVENTION

The invention will now be described in detail with reference to Figs. 1 to 6:

**Fig. 1** (A) Schematic representation of the ERhV1 genome and (B) comparison of the genomic structures of picornaviruses showing the predicted proteolytic cleavage pattern of the polyprotein. The lengths of individual regions are drawn approximately to scale. The dashed line represents the unsequenced region of the ERhV1 5'-NTR.

**Fig. 2a** Nucleotide and predicted amino acid sequence of the ERhV1 polyprotein. The nucleotide sequences of the 3'-NTR and part of the 5'-NTR are also shown. Numbering is from the first ATG codon that occurs in a context optimal for translational initiation (Kozak, 1989). A polypyrimidine tract upstream of the putative initiating ATG and the two pairs of in-frame ATG codons are underlined. The predicted proteolytic cleavage sites are indicated by arrows.

**Fig. 2b** Nucleotide sequence of the ERhV1 5'-nontranslated region. The polyC tract (dotted underline), polypyrimidine tract (underline) and potential initiation codons (double underline) are indicated. Predicted coding sequence is shown in bold type. Numbering is from the ATG considered most likely to be used for translation initiation.

**Fig. 3** Alignment of the predicted amino acid sequences of ERhV1.393/76 and FMDV.O1K polyprotein. Proteolytic cleavage sites, which are predicted in the case of ERhV1, are indicated by the arrows. Identical residues (*), highly conserved residues (:), and less conserved residues (.), are indicated.

**Fig. 4** Unrooted phylogenetic trees inferred using the picornavirus nucleotide sequences of (A) the complete polyprotein gene, (B) the polymerase gene and (C) VP1 gene of viruses representing the five recognised genera of the family *Picornaviridae*. The viruses used were: FMDV.A10, FMDV.O1K, FMDV.A12, FMDV.C3, FMDV.SAT3, EMCV, TMEV, Mengovirus, poliovirus 1.Mahoney (Polio 1), poliovirus 2.Sabin (Polio 2), poliovirus 3.Leon (Polio 2), coxsackievirus A9 (CV.A9), CV.B3, echovirus 22 (Echo 22), swine vesicular disease virus (SVDV), bovine enterovirus (BEV) hepatitis A virus (HAV) human rhinovirus 1B (HRV1B), HRV89 and HRV14.

Note: The branch lengths represent proportionate change only within each tree; they do not allow direct comparisons to be made between the three trees.

**Fig. 5(A)** Diagram outlining the strategy for nested, reverse transcription-polymerase chain reaction (RT-PCR) for the detection of ERhV genome. The genome structure of ERhV1 is shown schematically (top), and the first round PCR product (362bp), corresponding to VP1 and 2A regions, and the second round PCR product (210bp), corresponding to part of VP1, are represented as black lines.

(B) the sequence of specific oligonucleotide primers used for RT-PCR are shown. VP1R1 was used for the RT reaction.

**Fig. 6** Construction of ERhV1 expression plasmid for *E. coli* and baculovirus transfer vector for insect cells. The ERhV1 genome is shown (top) and oligonucleotide primers used to amplify P1.2A and 3C regions are depicted as

arrows. The P1.2A fragment and subsequently the P1.2A.3C fragment, obtained through the ligation of P1.2A and 3C, were cloned separately into the multiple cloning sites of the pET15b and pBacbluIII plasmid vectors to construct pET.P1.2A and pET.P1.2A.3C respectively for expression in *E. coli* and pBac.P1.2A and pBac.P1.2A.3C respectively for expression in insect cells.

The sequence of specific oligonucleotide primers used for the construction of expression plasmids are:

| | | | |
|---|---|---|---|
| VP4F | 5' | GCTGGATCCATGAGTGGCGGAGGTACATCCACT | 3' |
| R2A | 5' | GCTCTGCAGCAGGTCTGCTGATGCTTTGGA | 3' |
| 3CF | 5' | GCTCTGCAGATGATTAGGACTGAGACTGGTGT | 3' |
| 3CR | 5' | GCTGGATCCTTAGCCATAGTCAGGTTTGAA | 3' |

**Virus growth and purification**

ERhV1 strain 393/76 was isolated from a nasal swab taken from a thoroughbred horse in South Australia while it was being held in quarantine following importation from the United Kingdom. The mare had an acute, systemic febrile illness. The virus was passaged 14 times in equine fetal kidney (EFK) monolayer cell cultures and then once in Vero cells. ERhV1 virions were purified by a modification of the procedure described by Abraham and Colonno. Cells were harvested 48 hours after infection. The infected cells and supernatant fluid were frozen and thawed three times and clarified by centrifuging at 2,000 x g for 20 min at 4 C. Polyethylene glycol 6000 and NaCl were added to the supernatant to final concentrations of 7% and 380 mM, respectively, and the mixture was stirred overnight at 4 C. The precipitated virions were recovered by centrifuging at 10,000 x g for 15 min at 4 C and resuspended in 200-400 μl TNE buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA) containing 1% NP40. The suspension was clarified by centrifuging at 12,000 x g for 3 min before layering onto 15% to 45% (wt/vol) linear sucrose gradients (35 ml) in TNE buffer and centrifuging at 100,000 x g for 4 h at 4 C. Gradients were fractionated and the fractions analyzed by SDS-PAGE. Viral fractions were pooled, centrifuged at 200,000 x g for 2 h at 4 C, and the viral pellet was resuspended in a small volume of TNE buffer, cDNA

synthesis and cloning. Viral RNA was reverse transcribed using an oligo-dT primer (Amersham) or ERhV1 specific primers P1 (5'-ATCCAGCAAGCCGCTGTCCGGTTAC-3') and P5 (5'-CGAAGAGACACCTGCTTC-3'). Viral RNA was prepared as described in (1987) Anal-Biochem. 162, 156-159.

Viral RNA and 100 pmol of primer were mixed, boiled for 2 min and cooled at room temperature. First strand cDNA was synthesized using 200 U of Maloney murine leukemia virus reverse transcriptase (Promega) in the presence of 0.8 mM dNTPs and 30 U of human placental RNAse inhibitor (Pharmacia) in a reaction volume of 25 μl. Second strand cDNA was synthesized using a cDNA synthesis kit (Amersham). The cDNA fragments were ligated into pUC18, either as blunt ended fragments or after ligating BamH I adaptors (Pharmacia), and the lighted products used to transform *E. coli* strain DH5α (Stratagene). Colonies were selected by hybridization, initially with an [32P]-dCTP-labelled cDNA probe derived from reverse transcribed viral RNA, and subsequently with [32P]-dCTP-labelled cloned viral cDNA (16). The sequence between two cDNA clones was obtained using the oligonucleotide primers P6 (5'-TTCTGGTGGAGAAGTGACAGC-3') and P7 (5'-GTGAGCCAGCAACAATTGC-3') in a polymerase chain reaction (PCR; 17) using the polymerase, Vent Exo+ (New England Biolabs).

**DNA sequencing and analyses**

Double-stranded DNA was prepared using the alkaline lysis method and sequenced by dideoxy chain termination using modified T7 DNA polymerase (Pharmacia) and [35S]-dATP (Amersham). Sequence was read and analyzed using the GeneWorks software package (IntelliGenetics, Mountain View, CA). The GenBank database was searched using the FASTA searching and comparison program. The protein alignment shown in Fig. 3 was performed using the Genetics Computer Group, Inc. (Madison, Wisconsin, USA, 1994) GAP program with a gap creation penalty (GCP) of 3.0 and a gap extension penalty (GEP) of 0.1. The multiple alignments of nucleotide sequences were performed using ClustalW. For pairwise alignments the slow method was used with a GCP of 10 and a GEP of 0.1.

For multiple alignments a GCP of 10 and a GEP of 0.05 was used, with alignment of sequences which were more than 60% divergent delayed and using weighted transitions. Phylogenetic relationships were examined using the maximum likelihood method with the DNAML program of the Phylogeny Inference Package (Phylip) version 3.5c (1993. J. Felsenstein, Department of Genetics, University of Washington, Seattle). The model used allowed for unequal expected frequencies of the four nucleotides, with the frequencies determined empirically from those present in the sequences analysed, and unequal rates of transitions and transversions. A single rate of change was assumed for all sites. The program was allowed to perform global rearrangements to optimise the tree. Initial analyses were performed on polymerase sequences using a range of transition/transversion ratios to determine that which gave the maximal log likelihood. A ratio of 2.0 gave the maximal log likelihood and thus this ratio was used for all subsequent analyses of other sequences.

**Cloning And Sequencing Of The ERHV1 Genome**

Sixty seven overlapping cDNA clones and one PCR product clone were obtained and sequenced from both ends. The nucleotide in each position was determined at least twice, and 95% of the sequence was obtained by sequencing in both directions. The predicted genomic structure of ERhV1 was characteristic of picornaviruses, possessing one long open reading frame (ORF) flanked by 5'- and 3'-NTR's (Fig. 1).

The nucleotide and predicted amino acid sequences of the ERhV1 polyprotein are shown in Fig. 2a. Partial sequence of the 5'-NTR (433 bases) was also obtained Fig. 2b. There was a tract of 9 Cs at position -550 to -542. PolyC tracts of various lengths have been observed in similar locations in FMDV and EMCV. The actual length of the ERhV1 polyC tract is uncertain as these sequences are known to be unstable when propagated in *E. coli*. A 14 nucleotide polypyrimidine tract, which possessed the TTTC motif common to all picornaviruses, was present near the potential translation initiation codons. A region of 450 nucleotides upstream of the most likely initiation codon is predicted to contain an internal ribosome entry site (IRES). This region showed most

sequence identity (48-50%) with corresponding sequences in FMDV and EMCV. The 3'-NTR of ERhV1 was 102 nucleotides excluding the polyA tail (data not shown).

In picornaviruses, there are two factors that influence which ATG codon initiates translation, a requirement for the ATG to be located at the 3'-end of the IRES, and that this ATG occurs in a sequence optimal for initiating translation, that is, a purine at position -3 and a G in position +4. Two pairs of in-frame ATG codons were identified in the ERhV1 genome. The second ATG of the first pair is separated by 25 nucleotides from the beginning of the polypyrimidine tract (Fig. 2b), similar to the distance (25 to 27 nucleotides) found in the corresponding regions in FMDV and EMCV (24). The second ATG of each pair occurs in an optimal context. Therefore, the second ATG of the first pair is most likely to be the translation initiation codon but it is possible that translation is also initiated from the second optimal ATG, by a process of leaky scanning, or even from the other two, non-optimal ATG codons. The predicted ERhV1 coding sequence, beginning at the most likely initiation ATG, extended for 6,741 bases and would encode a polyprotein of 2,247 amino acids.

Alignment of the ERhV1 amino acid sequence with those of other picornaviruses showed that it was most similar to aphthoviruses and, to a lesser extent, to cardioviruses in all regions of the genome (data not shown). Fig. 3 shows a comparison of the predicted amino acid sequence of ERhV1 with that of FMDV.O1K. The two sequences were 40% identical. The more conserved regions include: the 3D/polymerase (50% identity), VP4 (49% identity) and some regions of the 2C protein. ERhV1 encoded a 2A protein of 16 amino acids, 14 of which were identical with those of FMDV 2A. ERhV1 possessed only one copy of the VPg sequence. This is in contrast to FMDV which has 3 tandemly repeated, non-identical VPg sequences (27-29).

Table 1 shows the proteolytic cleavage sites of ERhV1 predicted from the amino acid alignment (Fig. 3), and compares these with those of FMDV, EMCV and Theiler's murine encephalomyelitis virus (TMEV). Most of the ERhV1 cleavage sites could be assigned with reasonable confidence because of significant

amino acid similarity with FMDV in the regions flanking the predicted cleavage site; an exception was the 3A/3B cleavage site where there was less sequence similarity. As is the case with FMDV, the predicted ERhV1 3C protease cleavage sites were more variable than those of the cardioviruses, EMCV and TMEV.

Table 1. Comparison of the predicted proteolytic cleavage sites of the ERhV1 polyprotein with those of FMDV, EMCV and TMEV.

| Proteins | Cleavage sites* | | | |
|---|---|---|---|---|
| | ERhV1 | FMDV | EMCV | TMEV |
| Leader/1A(VP4) | S/G | K/G | Q/G | Q/G |
| 1A(VP4)/1B(VP2) | A/D | A/D | A/D | L/D |
| 1B(VP2)/1C(VP3) | E/A | E/G | Q/S | Q/S |
| 1C(VP3)/1D(VP1) | Q/V | E/T | Q/G | Q/G |
| 1D(VP1)/2A | T/N | L/N | E/S | E/N |
| 2A/2B | NPG/P | NPG/P | NPG/P | NPG/P |
| 2B/2C | Q/V | Q/L | Q/S | Q/G |
| 2C/3A | Q/S | Q/I | Q/G | Q/S |
| 3A/3B | Q/S | E/G | Q/G | Q/G |
| 3B/3C | E/T | E/S | Q/G | Q/G |
| 3C/3D | Q/G | E/G | Q/G | Q/G |

* Cleavage data from: FMDV.O1K (Forss *et al.* 1984), TMEV (Pevear *et al.* 1987) and EMCV (Palmenberg *et al.* 1984). The single amino acid code is used.

**Phylogenetic analyses**

A phylogenetic tree was derived from the nucleotide sequences of complete picornavirus polyproteins (Fig. 4a). Each branch of this tree was statistically, highly significant (P<0.01), with the 95% confidence limits ranging from ±7% to ±15% of branch lengths. ERhV1 was found to be most closely related to the aphthoviruses, although it was clear that ERhV1 was considerably more distant from individual members of this genus than the aphthoviruses were from each other. A phylogenetic tree was also derived from the nucleotide sequences of picornavirus polymerase genes (Fig. 4b). Each branch of this tree was statistically,

highly significant ($P<0.01$) with 95% confidence limits ranging from ±14% to ±38% of the branch lengths. Again, ERhV1 grouped with the aphthoviruses and the topology of the tree was the same as that obtained using data of the entire polyprotein (Fig. 4a). The VP1 nucleotide sequences were also similarly analyzed (Fig. 4c). Most branches were statistically, highly significant ($P<0.01$), although, that between the ERhV1 branch point and the branch point for the echovirus 22-hepatovirus cluster was less so ($P<0.05$). The 95% confidence limits of the branch lengths of this tree were considerably greater than for the other two trees, ranging from ±18% to ±69%. This tree did not group ERhV1 with the aphthoviruses. With the exception of bovine enterovirus (BEV), the tree had the same topology as those derived from the complete polyprotein and the polymerase sequences. It was also apparent that picornaviruses formed three clusters: enteroviruses-rhinoviruses, echovirus 22-hepatovirus and cardioviruses-aphthoviruses-ERhV1.

**(1) Diagnostic reagents**

**Oligonucleotide primers:** We have designed short oligonucleotide primers and used them in polymerase chain reactions (PCR) for the diagnosis of ERhV infected horses. Any of the ERhV nucleotide sequence may be used for the design primer sets for use as diagnostic reagents. They may be highly specific for ERhV1 or they may be designed to be more cross reactive so as to amplify single strand RNA template from other ERhV types e.g., ERhV 2, 3 and 4. As a specific example we have used the primer set shown in Fig. 5 to diagnose ERhV disease in several groups of seriously ill horses in circumstances in which, despite exhaustive efforts, we could not isolate the virus using conventional cell culture procedures. We now consider ERhV a very under reported disease simply because, most of the time, nasal samples collected from horses experiencing severe, systemic clinical disease because of ERhV infection do not yield the virus in cell culture. In one particular group of horses, we detected the presence of ERhV by PCR and confirmed that the horses were both actively infected and seriously ill with ERhV by use of paired serum samples which showed that there was a concomitant rise in

ERhV1 serum neutralising antibody. Vigorous attempts to isolate the virus in cell cultures yielded negative results.

**Oligonucleotide probes:** Virus specific oligonucleotides are used as probes to detect the presence of the virus in infected samples from diseased horses and other animals. This may be especially important given the systemic nature of the illness i.e., it is a foot-and-mouth-like, generalized disease with virus distributed throughout the body in many organs and tissues; it is not just a simple "common cold-like" illness as the name rhinovirus implies The significance of the sequence in moving the virus out of the *Rhinovirus* genus and into a new genus proposed to be called "*Equirhinovirus*" in the *Picornaviridae* family does not represent merely a taxonomic change but represents a paradigm shift in how ERhV1 and related viruses must now be regarded as pathogens for the horse and other animal species.

**Diagnostic antigens:** Individual virion proteins, in particular VP1, VP2 and VP3, can be expressed in any one of a number of heterologous expression systems to provide antigens to detect specific antibody to ERhV1 present in blood. Such expression systems, which are well established for *E. coli*, yeast, vaccinia virus and baculovirus, allow for the production of large quantities of protein to a high degree of purity. The expressed virion proteins may be used in simple immunoassays, such as ELISA, to detect ERhV1 specific antibody. Virion proteins expressed in this way also serve as effective vaccines against ERhV1 disease.

**(2) Vaccines**

**Production of virus like particles (VLPs):** We have used the sequence information to construct recombinant plasmids containing the P1-2A-3C region of the genome (see Fig. 1a and Fig. 6). These plasmid constructions are of course critically dependent on the ERhV1 sequence that has been determined although the strategy that we are adopting, in general, is similar to that described in J. Virol 66, 4557-4564. Some early plasmid constructions have been inserted into *E. coli* and baculovirus expression systems based on prior art with similar viruses such as poliomyelitis of humans and foot-and-mouth disease virus of cattle and other cloven hoofed animals. The RT PCR double stranded DNA of the P1-2A-3C region of the ERhV1 genome is transcribed, within the transformed *E. coli* or insect cell for

baculovirus, into messenger RNA as a single transcript which is then translated into a mini polyprotein. The 3C protease activity results in the cleavage of the mini polyprotein into its constituent parts namely 1A (VP4), 1B (VP2), 1C (VP3) and 1D(VP1), 2A and 3C (see Fig. 1a and Fig. 6) and that the VP component parts then self assemble into VLPs i.e., virus particles that lack nucleic acid and are therefore non infectious i.e., are unable to cause disease. Two important applications of ERhV VLPs are as follows:

(a) The VLPs are very useful as highly effective, safe, high antigen-mass vaccines for the control ERhV1 disease. If ERhV1 disease is confirmed, as we believe to be the case, as significant and responsible for much hither to undiagnosed illness that results in many lost training days, many expensive treatments, much serious illness because of secondary infections following on the primary ERhV1 infection, and much poor performance, then the utility of the vaccine based on the VLPs that are the subject of this invention will be very great and likely to have world-wide application.

With improved methods for the diagnosis of ERhV1 infection such as by PCR and ELISA as described herein, it is likely that other members of the proposed new *Equirhinovirus* genus within the family *Picornaviridae* including for example ERhV2, ERhV3, may be similarly diagnosed. Indeed suitably selected PCR primer sets based on the ERhV1 sequence could be used to detect these other equine rhinoviruses. The sequencing of these genomes could provide a basis for their specific diagnosis. It is also evident that the construction of VLP's based on expression plasmids similar to those described herein for ERhV1, could be readily adapted to these other equine rhinoviruses leading for example to production of combined ERhV vaccines to cover all antigenic types as may be extant or as may emerge by antigenic variation, as is very much a part of the biology of FMDV, in the future. Polyvalent VLP vaccines incorporating a range of ERhV antigenic types are obvious extensions based on the work described herein.

(b) ERhV VLPs can be used as a delivery vector that will provide not only protection against ERhV disease but will be used to deliver other therapeutic and useful substances to the horses following administration by parenteral or other

routes. Such delivery vectors can be produced by inserting into, for example the P1 region at some appropriate site, double stranded DNA coding for antigenic epitopes of other virus and infectious agents of horse as well as epitopes derived from other non infectious sources for example reproductive hormones.

## ERhV1 DIAGNOSTIC TESTS

For the detection of ERhV1 antibodies in infected or vaccinated horses various standard tests can be used. VLP's may be used in such tests for example in an ELISA test for antibody.

Other diagnostic tests based on recombinant antigens derived from the ERhV1 sequence can be devised along similar lines to those reported for FMDV in which the absence of protein 2C from clarified inactivated whole virus FMD, FMDV or FMDV VLP vaccines maybe used as the basis for distinguishing infected from vaccinated animals where the vaccine is a non-replicating form of ERhV1 or a deletion mutant of ERhV1 in which a particular non-structural protein gene has been deleted. Precedent for this comes from studies of FMDV as reported in for example Lubroth, Grubman, Burrage, Newman & Brown, 1996, Absence of protein 2C from clarified foot-and-mouth disease virus vaccines provides the basis for distinguishing convalescent from vaccinated animals, Vaccine 14(5), 419-427.

## PREPARATION AND USE OF VIRUS-LIKE PARTICLES AND OTHER PROTEINS BASED ON ERhV1 SEQUENCE

From the sequence of ERhV1 it is possible to clone certain segments of the viral genome into a variety of vectors for expression in a variety of different expression systems. There is a straight forward and strong literature for FMDV that provides a very clear precedent for what can be done for ERhV1. Examples include the expression of FMDV P1-2A in a baculovirus (Abrams CC & Belsham GJ, 1994, The antigenicity of foot-and-mouth disease virus P1-2A polyprotein and empty capsids produced in vaccinia virus and baculovirus expression systems. In VIIth Meeting of the European Study Group on the Molecular Biology of Picornaviruses, 6-11 August 1994, Korpilampi, Finland) or vaccina virus syst ms (Abrams CC, King AMQ & Belsham GJ, 1995, Assembly of foot-and-mouth

disease virus empty capsides synthesized by a vaccinia virus expression system. Journal of General Virology 76:3089-3098) to obtain VLPs or viral proteins. We have prepared similar plasmids in which P1-2A, P1-2A-3C and these two sequences in a myristolated form have been inserted into p fastbac 1 baculovirus vector (Gibco/BRL) and into a PET vector (Novogene) for expression in insect cells and *E.coli* respectively.

These expressed products either as protein antigens or as VLPs, have utility as the basis for diagnostic tests or vaccines.

Accordingly, such references are herein incorporated in support of the full description and enablement of the invention where the disclosed methods of preparing diagnostics, vaccines, vectors, host systems and kits are fully described and applicable to the like aspects of the current invention.

**(3) Applications in human medicine:**

ERhV is also a human pathogen. We have unpublished data to confirm that humans have serum neutralising antibody to ERhV1 that is indicative of infection. One of the laboratory workers concerned with the conduct of the sequencing and who handled infectious virus has specific antibody in high amounts (serum neutralising antibody titre 1 : 640 to ERhV1). We are currently extending these studies and anticipate finding a significant incidence of infection in humans world wide particularly among those humans who work with horses. The improved diagnostic methods outlined above, perhaps also the vaccine, are expected to have application in human medicine.

## CLAIMS:

1. A substantially pure nucleotide sequence for ERhV1 being:

CCGTCAAGCC CGTTGCCTGT ATAGCCAGGT AACCGGACAG CGGCTTGCTG GATTTTCCCG -375
GTGCCATTGC TCTGGATGGT GTCACCAAGC TGACAAATGC GGAGTGAACC TCACAAAGCG -315
ACACGCCTGT GGTAGCGCTG CCCAAAAGGG AGCGGAACTC CCCGCCGAGG CGGTCCTCTC -255
TGGCCAAAAG CCCAGCGTTG ATAGCGCCTT TTGGGATGCA GGAACCCCAC CTGCCAGGTG -195
TGAAGTGGAG TGAGCGGATC TCCAATTTGG TCTGTTCTGA ACTACACCAT TTACTGCTGT -135
GAAGAATGCC CTGGAGGCAA GCTGGTTACA GCCCTGACCA GGCCCTGCCC GTGACTCTCG -75
ACCGGCGCAG GGTCAAAAAT TGTCTAAGCA GCAGCAGGAA CGCGGGAGCG TTTCTTTTCC -15
TTTTGTACTG ACATGATGGC GGCGTCTAAG GTGTATAGAG TTTGCGAGCA GACTCTGCTG 45
GCAGGTGCCG TTCGCATGAT GGACAAATTC TTGCAAAAGA GAACTGTTTT TGTCCCCCAT 105
CTTGACAAAA CAATTCGTTT GACTGGACTC CACAATTATG ACAATACTTG CTGGTTGAAT 165
GCCTTGACAC AACTGACACA GATTCTTGGA ATTCGGCTTT TTGATGAACA CTTCGGCAAT 225
AGAGGTCTGT TCACTCGGAA AACAATTGAT TGGGTGAGTG ACCAGACTGG TATAAAAGAT 285
CTAAAATCAG GAGCACCGCC ACTCGTGGTG GTGTACAAAC TGTGGCAACA TGGACACTTG 345
GATGTCGGTA CGATGGAGAA ACCCCGGTCG ATTACTCTAT GGTCTGGCCC CAAAGTGTGT 405
CTTTCTGATT TCTGGGCCTG TGTTTCGGCA AAACCGGGAC ATGCAGTATT CTACCTTCTC 465
ACAAGCGAGG GTTGGATCTG TGTTGATGAC AAGAAAATAT ACCCAGAAAC ACCCAAAACA 525
GAGGATGTAC TTGTTTTTGC GCCCTATGAC TTTGAGTCAC TGGGCAAGGA CCCACCAAAG 585
CTACACCAGA GATATGAAAA AGCATTTGAG CTCAGTGGCG GAGGTACATC CACTCCAACA 645
ACTGGCAACC AAAACATGTC CGGAAACAGT GGTTCAATTG TTCAAAATTT TTACATGCAA 705
CAGTACCAGA ATTCAATTGA CGCAGACCTG GGAGACAATG TGATTAGCCC TGAAGGCCAG 765
GGCAGCAACA CTAGTAGTTC AACCTCATCA AGCCAATCCT CTGGCTTGGG CGGGTGGTTC 825
TCTAGTTTGC TGAACCTTGG AACAAAACTA CTGGCTGACA AGAAGACAGA AGAGACTACA 885
AACATTGAAG ACAGAATTGA AACAACAGTG GTTGGAGTCA CTATTATTAA TTCACAAGGA 945
TCTGTTGGAA CAACCTACTG TTACTCCAAA CCGGATGGTA GACCACCATC CACAGTGTCA 1005
GACCCAGTTA CCAGACTTGG ACCCACGCTT TCCAGGCACT ACACATTTAA GGTAGGTGAG 1065
TGGCCCCATT CTCAATCACA TGGTCACGCA TGGATCTGTC CGTTGCCAGG TGACAAACTC 1125
AAGAAGATGG GCAGTTTTCA TGAGGTTGTC AAAGCCCACC ACCTGGTCAA GAACGGCTGG 1185
GATGTGGTTG TGCAGGTGAA TCCCTCATTT GCTCACTCCG GGCCGCTGTG TGTAGCAGCA 1245
GTGCCGGAGT ACGAACACAC ACATGAGAAA GCACTCAAGT GGTCTGAGCT TGAGGAACCA 1305
GCTTACACAT ACCAACAACT TTCAGTTTTT CCCCACCAGT TGCTAAATTT GAGGACAAAT 1365
TCATCAGTGC ATTTGGTGAT GCCCTACATT GGGCCAGGCC AACCAACAAA TCTGACTTTG 1425
CACAACCCGT GGACCATTGT TATTTTAATT TTGTCTGAAT TGACAGGACC TGGCCAAACT 1485
GTGCCTGTGA CCATGTCGGT GGCTCCCATC GATGCAATGG TTAATGGGCC TCTTCCAAAT 1545
CCAGAGGCAC CGATTAGAGT GGTGTCTGTG CCTGAATCAG ATTCTTTTAT GTCTTCAGTA 1605
CCTGATAATT CGACTCCACT ATACCCCAAG GTTGTGGTCC CACCGCGCCA AGTTCCTGGC 1665
CGGTTTACAA ATTTCATTGA TGTGGCAAAA CAGACATATT CATTTTGTTC CATTTCTGGA 1725
AAACCTTATT TTGAGGTTAC CAACACCTCT GGGGACGAGC CACTGTTTCA GATGGATGTG 1785
TCGCTCAGTG CGGCAGAGCT ACATGGCACT TACGTAGCTA GTTTGTCATC ATTTTTTGCA 1845
CAGTACAGAG GCTCACTTAA TTTCAACTTT ATTTTCACTG GTGCAGCAGC CACTAAGGCA 1905
AAGTTTCTGG TTGCTTTTGT GCCTCCCCAC AGTGCAGCGC CCAAAACGCG CGATGAAGCA 1965
ATGGCGTGCA TCCATGCCGT GTGGGATGTT GGCTTGAACT CAGCTTTTTC TTTTAATGTA 2025
CCTTATCCCT CCCCTGCTGA CTTCATGGCC GTTTATTCTG CGGAACGGAC GGTTGTGAAT 2085
GTCTCTGGAT GGCTTCAAGT TTATGCACTA ACAGCTCTAA CTTCAACTGA CATTGCCGTG 2145
AACAGTAAAG GCCGTGTGCT GGTTGCTGTT TCCGCCGGCC CAGACTTCTC CCTTCGTCAC 2205

CCGGCGGACC TGCCCGACAA GCAGGTTACC AATGTGGGAG AGGATGGTGA ACCCGGTGAG 2265
ACAGAGCCTC GTCATGCTTT GTCACCCGTG GACATGCACG TGCACACAGA TGTCAGTTTC 2325
TTGCTTGACC GGTTCTTTGA TGTTGAGACA CTTGAGCTTT CAAATTTGAC AGGTTCTCCT 2385
GCCACACATG TTCTGGATCC GTTTGGCTCG ACTGCCCAAC TGGCTTGGGC ACGTCTGCTA 2445
AACACTTGCA CCTACTTCTT TTCTGATTTG GAATTGTCAA TCCAGTTTAA ATTTACCACC 2505
ACTCCGTCCT CTGTTGGAGA GGGCTTTGTG TGGGTGAAGT GGCTCCCTGT TGGAGCACCA 2565
ACCAAGACCA CAGATGCTTG GCAGTTAGAA GGAGGTGGAA ATTCAGTTAG AATTCAAAAA 2625
TTGGCCGTTG CAGGGATGTG CCCCACTGTT GTGTTCAAGA TTGCAGGCTC CCGTTCACAA 2685
GCCTGTGCTT CAGCGTTGCC ATATACATCA ATGTGGCGTG TTGTGCCAGT CTTTTACAAT 2745
GGCTGGGGTG CACCTACCAA AGAAAAGGCA ACCTACAATT GGCTTCCTGG TGCACACTTT 2805
GGTTCCATCT TGCTGACTTC TGATGCGCAT GATAAAGGAG GGTGCTACTT GCGGTATGCT 2865
TTCCGCGCGC CAGCGATGTA TTGCCCTCGA CCCATTCCGC CGGCTTTTAC GCGTCCAGCG 2925
GACAAAACCA GACATAAATT TCCCACTAAC ATCAACAAAC AGTGTACTAA TTACTCTCTC 2985
CTCAAATTGG CTGGAGATGT TGAGAGCAAC CCTGGCCCCA CTATTTTTTC CAAAGCATCA 3045
GCAGACCTGA ATGCCTTGTC AACGTCGCTA GGTGAATTGA CTGGCATGCT AAAAGATCTT 3105
AAAGCCAAGG CAGAAACTTA TTCCCCGTTT TACAAAATGG CCAAAATGCT TTTCAAACTT 3165
GCAACACTAG CTGTGGCAGC TATGAGGACA AAGGACCCAG TAGTGGTGGT TATGTTGATT 3225
GCTGATTTCG GATTGGAGGT CTTTGACACT GGGTTTTTCT TTTCCTACTT TCAAGAGAAG 3285
TTGCAGCCTT ATATGAAAAC TATTCCTGGT AAGATTTCTG ATTTGGTCAC TGATGCGGCT 3345
ACGGCTGCCG CCCAAATTCC AAAGGGAGTG TATTCTTTTG TGTCGTCATT TTTCGAAACG 3405
CCTGAAGGAG TGGTTGAGAA GCAGGTGTCT CTTCGGACAG TGAATGACAT ATTTGCTTTG 3465
CTTAAAAATT CTGATTGGTT CATAAAGACT CTTGTTGCCC TCAAGAAATG GCTGACATCC 3525
TGGTTTGCTC AAGAACAACA GGCAGATGAT GCGCTCTATT CAGAATTGGA AAAATATCCC 3585
TTGTACAAGT TAAAATTGAA GGAACCTGAT ACTCAAGAGG AAGCGCGCCA GTGGTTTAAA 3645
GACATGCAGC AGCGTGCTCT CGCTGTGAAG GACAAAGGTC TCTTTTCCCT CCTGCAAATT 3705
CCATTAGTTA ACTTGCCCCA GAGCCGTCCA GAGCCCGTTG TATGCGTCCT TCGGGGCGCA 3765
TCAGGGCAAG GCAAATCTTA TTTGGCAAAT CTGATGGCTC AAGCAATTTC GCTTCTCTTG 3825
GTTGGCAAGC AGGACAGTGT GTGGAGTTGT CCTCCTGACC CCACATATTT TGATGGCTAT 3885
AACGGACAGG CTGTGGTGAT TATGGATGCA TTGGGCCAGG ATCCGAATGG TGCTGACTTT 3945
AAATATTTTT GCCAGATGGT CTCTACAACA GCTTTTGTAC CACCTATGGC CCATTTGGAT 4005
GATAAAGGCA TTCCATTTAC TTCTCCTGTT GTTATTTGTA CTACAAATTT GCATTCATCT 4065
TTTACCCCTA TTACTGTTTC TTGTCCTGAA GCTCTTAAGA GGAGGTTTCG GTTTGATGTG 4125
ACGGTGTCCG CTAAACCGGG CTTTGTGCGC ACTGTTGGTT CAAACCAGCT TTTGAATCTC 4185
CCACTTGCTC TTAAGCCAGC TGGTCTTCCC CCACACCCTA TCTTTGAAAA TGACATGCCC 4245
ATTATAAATG GGCAGGCTGT TAAATTGGCT CTTTCTGGTG GAGAAGTGAC AGCTTTTGAG 4305
CTTATTGAGA TGATACTGTC AGAAGTTCAA AACAGACAAG ACACACACAA AATGCCCATT 4365
TTTAAACAAT CATGGTCTGA TTTGTTCAGA AAGTGTACAA CTGATGAGGA ACAGAAAATG 4425
TTGCAGTTTT TAATTGACAA TAAAGATTCA GAAATTCTCA GGGCGTTTGT TTCAGAACGC 4485
TCCATTTTAC TACATGAAGA GTATCTTAAA TGGGAGTCAT ATATGACCAG GAGAGCCAAG 4545
TTTCACCGCC TGGCTGCTGA TTTTGCTATG TTTCTATCCA TTCTTACTTC ACTGATTGTT 4605
ATTTTTTGTT TAGTTTATTC TATGTATCAA CTTTTTAAGA CCCCTGACGA GCAATCAGCT 4665
TATGATCCTT CAACTAAGCC AAAACCAAAG ACCCAGGAAG TGAAAACACT GAAGATTAGG 4725
ACTGAGACTG GTGTACCAGC AACTGACTTG CAACAATCCA TCATGAAAAA TGTTCAGCCA 4785
ATTGAGCTTT ACCTTGACAA TGAATTGGTT ACTGACTGCT CTGCCTTGGG TGTTTATGAC 4845
AATTCATATT TGGTGCCCCT TCATTTGTTT GAATTTGATT TTGATACCAT TGTGCTTGGT 4905
GGACGTCATT ACAAGAAAGC TGAGTGTGAG AAGGTAGAGT TTGAGCTTGA AGTGAATGGA 4965
GACGTGGTGT CATCAGATGC GTGTCTACTT CGAGTGTCAT CGGGGCCTAA AGTTAGAAAT 5025

```
ATTGTTCATC TTTTTACAAA TGAAATTGAA TTGAAGAAAA TGACCCAAGT GACAGGAATC   5085
ATGAATTCAC CACACCAGGC ACGCACTGTG TTTTTTGGCA GTTTTTTGAC AGTGAGGAAG   5145
TCCATCTTAA CATCGGATGG GACTGTAATG CCCAATGTTT TGTCCTATGC CGCTCAGACC   5205
TCGCGTGGGT ATTGTGGCGC TGCAATTGTT GCTGGCTCAC CTGCCCGCAT AATTGGTATC   5265
CATTCAGCTG GCACTGGATC TGTTGCATTT TGCTCCCTGG TGTCCAGAGA CGCGCTGGAG   5325
CAACTCTGGC CCCAGAAACA GGGCAACGTT AGTCGCCTTG ATGACGATGT GAGGGTGTCT   5385
GTTCCGCGCC GCTCCAAATT GGTGAAATCA TTGGCTTACC CCATTTTCAA ACCTGACTAT   5445
GGCCCAGCGC CACTCTCTCA ATTTGACAAG CGCCTGTCAG ACGGCGTGAA GCTGGATGAA   5505
GTGGTTTTTG CTAAACATAC TGGAGACAAG GAGATTTCCG CACAGGACCA GAAATGGCTC   5565
TTGCGTGCGG CGCATGTATA CGCCCAGAAG GTTTTCTCCC GGATTGGATT TGACAACCAG   5625
GCTTTGACTG AAAAAGAGGC CATTTGTGGC ATTCCTGGCC TTGACAAGAT GGAGCAGGAC   5685
ACCGCTCCCG GGCTGCCCTA TGCTCAGCAA AATAAGAGAA GGAAAGACAT CTGTGATTTT   5745
GAAGAGGGCC GGCTGAAGGG CGCCGAACTC CAAAAGGACA GATTTATGGC TGGTGACTAC   5805
TCTAATTTGG TCTATCAATC ATTTTTGAAA GATGAGATCC GCCCACTTGA GAAAGTTAGG   5865
GCTGGAAAGA CCCGCCTGAT TGACGTGCCG CCGATGCCCC ATGTGGTGGT TGGTAGGCAG   5925
CTCTTGGGCC GGTTTGTGGC AAAATTTCAT GAAGCAAATG GATTTGACAT TGGCTCAGCC   5985
ATTGGATGTG ACCCAGATGT GGACTGGACT CGGTTTGGCC TCGAGTTGGA GCGTTTCAGG   6045
TATGTATATG CCTGTGACTA CTCACGGTTC GATGCCAACC ATGCAGCTGA TGCAATGAGA   6105
GTTGTGCTTA ACTACTTTTT CTCTGAGGAC CACGGTTTCG ACCCTGGTGT GCCTGCTTTT   6165
ATTGAGTCAC TGGTTGATTC AGTGCATGCC TATGAAGAGA AAAGGTATAA CATCTACGGT   6225
GGCTTGCCAT CCGGGTGTTC CTGCACATCA ATTTTGAATA CCATCTTGAA CAATGTTTAC   6285
ATTCTTGCAG CTATGATGAA GGCTTATGAG AATTTTGAGC CAGATGACAT TCAGGTCATT   6345
TGCTATGGGG ACGACTGCCT CATTGCTTCT GATTTTGAAA TTGATTTCCA ACAACTGGTG   6405
CCTGTCTTTT CTAGTTTTGG ACAGGTAATA ACTACAGCTG ACAAGACTGA TTTTTTTAAA   6465
CTGACAACGC TTTCGGAGGT GACCTTCCTT AAGCGCGCTT TTGTTCTGAC GGCCTTTTAC   6525
AAGCCAGTGA TGGATGTGAA GACCCTTGAA GCAATCTTAA GCTTTGTTCG CCCAGGCACA   6585
CAGGCTGAAA AGCTCCTGTC CGTGGCGCAG TTGGCAGGCC ACTGCGAACC GGAGCAGTAT   6645
GAGCGCCTGT TTGAGCCCTT TGCTGGGATG TATTTCGTCC CTACTTGGCG ACTTGCGCCT   6705
GCAGTGGTTG ATGAAGCTTG GATGCTAAAT TCTTTTTGAC TTTGTTTTTC TTTGTTTTCT   6765
TTTAGGCTTT TAAGGTGTTA AGTTTAAAGG TTAAGAGTTT TTAGAAGTTA AGATAGAGTT   6825
TAGTTTTTAG TTTTGAGC-poly(A)
```

as disclosed in Fig. 2 and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants, degeneracy equivalents and deletion mutants thereof.

2. A substantially pure amino acid sequence being:

M A A S K V Y R V C E Q T L L A G A V R M M D K F
L Q K R T V F V P H L D K T I R L T G L H N Y D N
T C W L N A L T Q L T Q I L G I R L F D E H F G N
R G L F T R K T I D W V S D Q T G I K D L K S G A
P P L V V V Y K L W Q H G H L D V G T M E K P R S
I T L W S G P K V C L S D F W A C V S A K P G H A
V F Y L L T S E G W I C V D D K K I Y P E T P K T
E D V L V F A P Y D F E S L G K D P P K L H Q R Y

L ↓ VP4

E K A F E L S G G G T S T P T T G N Q N M S G N S
G S I V Q N F Y M Q Q Y Q N S I D A D L G D N V I
S P E G Q G S N T S S S T S S S Q S S G L G G W F

VP4 ↓ VP2

S S L L N L G T K L L A D K K T E E T T N I E D R
I E T T V V G V T I I N S Q G S V G T T Y C Y S K
P D G R P P S T V S D P V T R L G P T L S R H Y T
F K V G E W P H S Q S H G H A W I C P L P G D K L
K K M G S F H E V V K A H H L V K N G W D V V V Q
V N P S F A H S G P L C V A A V P E Y E H T H E K
A L K W S E L E E P A Y T Y Q Q L S V F P H Q L L
N L R T N S S V H L V M P Y I G P G Q P T N L T L
H N P W T I V I L I L S E L T G P G Q T V P V T M

VP2 ↓ VP3

S V A P I D A M V N G P L P N P E A P I R V V S V
P E S D S F M S S V P D N S T P L Y P K V V V P P
R Q V P G R F T N F I D V A K Q T Y S F C S I S G
K P Y F E V T N T S G D E P L F Q M D V S L S A A
E L H G T Y V A S L S S F F A Q Y R G S L N F N F
I F T G A A A T K A K F L V A F V P P H S A A P K
T R D E A M A C I H A V W D V G L N S A F S F N V
P Y P S P A D F M A V Y S A E R T V V N V S G W L
Q V Y A L T A L T S T D I A V N S K G R V L V A V

VP3 ↓ VP1

S A G P D F S L R H P A D L P D K Q V T N V G E D

G E P G E T E P R H A L S P V D M H V H T D V S F
L L D R F F D V E T L E L S N L T G S P A T H V L
D P F G S T A Q L A W A R L L N T C T Y F F S D L
E L S I Q F K F T T T P S S V G E G F V W V K W L
P V G A P T K T T D A W Q L E G G G N S V R I Q K
L A V A G M C P T V V F K I A G S R S Q A C A S A
L P Y T S M W R V V P V F Y N G W G A P T K E K A
T Y N W L P G A H F G S I L L T S D A H D K G G C
Y L R Y A F R A P A M Y C P R P I P P A F T R P A
VP1 ↓ 2A
D K T R H K F P T N I N K Q C T N Y S L L K L A G
2A ↓ 2B
D V E S N P G P T I F S K A S A D L N A L S T S L
G E L T G M L K D L K A K A E T Y S P F Y K M A K
M L F K L A T L A V A A M R T K D P V V V V M L I
A D F G L E V F D T G F F F S Y F Q E K L Q P Y M
K T I P G K I S D L V T D A A T A A A Q I P K G V
2B ↓ 2C
Y S F V S S F F E T P E G V V E K Q V S L R T V N
D I F A L L K N S D W F I K T L V A L K K W L T S
W F A Q E Q Q A D D A L Y S E L E K Y P L Y K L K
L K E P D T Q E E A R Q W F K D M Q Q R A L A V K
D K G L F S L L Q I P L V N L P Q S R P E P V V C
V L R G A S G Q G K S Y L A N L M A Q A I S L L L
V G K Q D S V W S C P P D P T Y F D G Y N G Q A V
V I M D A L G Q D P N G A D F K Y F C Q M V S T T
A F V P P M A H L D D K G I P F T S P V V I C T T
N L H S S F T P I T V S C P E A L K R R F R F D V
T V S A K P G F V R T V G S N Q L L N L P L A L K
P A G L P P H P I F E N D M P I I N G Q A V K L A
L S G G E V T A F E L I E M I L S E V Q N R Q D T
2C ↓ 3A
H K M P I F K Q S W S D L F R K C T T D E E Q K M
L Q F L I D N K D S E I L R A F V S E R S I L L H
E E Y L K W E S Y M T R R A K F H R L A A D F A M

F L S I L T S L I V I F C L V Y S M Y Q L F K T P

3A ♦ 3B

D E Q S A Y D P S T K P K P K T Q E V K T L K I R

3B ♦ 3C

T E T G V P A T D L Q Q S I M K N V Q P I E L Y L
D N E L V T D C S A L G V Y D N S Y L V P L H L F
E F D F D T I V L G G R H Y K K A E C E K V E F E
L E V N G D V V S S D A C L L R V S S G P K V R N
I V H L F T N E I E L K K M T Q V T G I M N S P H
Q A R T V F F G S F L T V R K S I L T S D G T V M
P N V L S Y A A Q T S R G Y C G A A I V A G S P A
R I I G I H S A G T G S V A F C S L V S R D A L E

3C ♦ 3D

Q L W P Q K Q G N V S R L D D D V R V S V P R R S
K L V K S L A Y P I F K P D Y G P A P L S Q F D K
R L S D G V K L D E V V F A K H T G D K E I S A Q
D Q K W L L R A A H V Y A Q K V F S R I G F D N Q
A L T E K E A I C G I P G L D K M E Q D T A P G L
P Y A Q Q N K R R K D I C D F E E G R L K G A E L
Q K D R F M A G D Y S N L V Y Q S F L K D E I R P
L E K V R A G K T R L I D V P P M P H V V V G R Q
L L G R F V A K F H E A N G F D I G S A I G C D P
D V D W T R F G L E L E R F R Y V Y A C D Y S R F
D A N H A A D A M R V V L N Y F F S E D H G F D P
G V P A F I E S L V D S V H A Y E E K R Y N I Y G
G L P S G C S C T S I L N T I L N N V Y I L A A M
M K A Y E N F E P D D I Q V I C Y G D D C L I A S
D F E I D F Q Q L V P V F S S F G Q V I T T A D K
T D F F K L T T L S E V T F L K R A F V L T A F Y
K P V M D V K T L E A I L S F V R P G T Q A E K L
L S V A Q L A G H C E P E Q Y E R L F E P F A G M

3D

Y F V P T W R L A P A V V D E A W M L N S F

3. A protein or virus like particle incorporating VP1. derived from ERhV1 and having the following amino acid sequence:

V T N V G E D G E P G E T E P R H A L S P V D M H
V H T D V S F L L D R F F D V E T L E L S N L T G
S P A T H V L D P F G S T A Q L A W A R L L N T C
T Y F F S D L E L S I Q F K F T T T P S S V G E G
F V W V K W L P V G A P T K T T D A W Q L E G G G
N S V R I Q K L A V A G M C P T V V F K I A G S R
S Q A C A S A L P Y T S M W R V V P V F Y N G W G
A P T K E K A T Y N W L P G A H F G S I L L T S D
A H D K G G C Y L R Y A F R A P A M Y C P R P I P
P A F T R P A D K T R H K F P T N I N K Q C T

4. A protein or virus like particle incorporating VP2, derived from ERhV1 and having the following amino acid sequence:

D K K T E E T T N I E D R I E T T V V G V T I I N
S Q G S V G T T Y C Y S K P D G R P P S T V S D P
V T R L G P T L S R H Y T F K V G E W P H S Q S H
G H A W I C P L P G D K L K K M G S F H E V V K A
H H L V K N G W D V V V Q V N P S F A H S G P L C
V A A V P E Y E H T H E K A L K W S E L E E P A Y
T Y Q Q L S V F P H Q L L N L R T N S S V H L V M
P Y I G P G Q P T N L T L H N P W T I V I L I L S
E L T G P G Q T V P V T M S V A P I D A M V N G P
L P N P E

5. A protein or virus like particle incorporating VP3, derived from ERhV1 and having the following amino acid sequence:

A P I R V V S V P E S D S F M S S V P D N S T P L
Y P K V V V P P R Q V P G R F T N F I D V A K Q T
Y S F C S I S G K P Y F E V T N T S G D E P L F Q
M D V S L S A A E L H G T Y V A S L S S F F A Q Y
R G S L N F N F I F T G A A A T K A K F L V A F V
P P H S A A P K T R D E A M A C I H A V W D V G L
N S A F S F N V P Y P S P A D F M A V Y S A E R T
V V N V S G W L Q V Y A L T A L T S T D I A V N S
K G R V L V A V S A G P D F S L R H P A D L P D K
Q

6. A protein or virus like particle incorporating VP4, derived from ERhV1 and having the following amino acid sequence:

G G G T S T P T T G N Q N M S G N S G S I V Q N F
Y M Q Q Y Q N S I D A D L G D N V I S P E G Q G S
N T S S S T S S S Q S S G L G G W F S S L L N L G
T K L L A

7. A substantially pure nucleotide sequence for VP1 being:

GTTACCAATG TGGGAGAGGA TGGTGAACCC GGTGAGACAG AGCCTCGTCA TGCTTTGTCA
CCCGTGGACA TGCACGTGCA CACAGATGTC AGTTTCTTGC TTGACCGGTT CTTTGATGTT
GAGACACTTG AGCTTTCAAA TTTGACAGGT TCTCCTGCCA CACATGTTCT GGATCCGTTT
GGCTCGACTG CCCAACTGGC TTGGGCACGT CTGCTAAACA CTTGCACCTA CTTCTTTTCT
GATTTGGAAT TGTCAATCCA GTTTAAATTT ACCACCACTC CGTCCTCTGT TGGAGAGGGC
TTTGTGTGGG TGAAGTGGCT CCCTGTTGGA GCACCAACCA AGACCACAGA TGCTTGGCAG
TTAGAAGGAG GTGGAAATTC AGTTAGAATT CAAAAATTGG CCGTTGCAGG GATGTGCCCC
ACTGTTGTGT TCAAGATTGC AGGCTCCCGT TCACAAGCCT GTGCTTCAGC GTTGCCATAT
ACATCAATGT GGCGTGTTGT GCCAGTCTTT TACAATGGCT GGGGTGCACC TACCAAAGAA
AAGGCAACCT ACAATTGGCT TCCTGGTGCA CACTTTGGTT CCATCTTGCT GACTTCTGAT
GCGCATGATA AAGGAGGGTG CTACTTGCGG TATGCTTTCC GCGCGCCAGC GATGTATTGC
CCTCGACCCA TTCCGCCGGC TTTTACGCGT CCAGCGGACA AAACCAGACA TAAATTTCCC
ACTAACATCA ACAAACAGTG TACT

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

8. A substantially pure nucleotide sequence for VP2 being:

GACAAGAAGA CAGAAGAGAC TACAAACATT GAAGACAGAA TTGAAACAAC AGTGGTTGGA
GTCACTATTA TTAATTCACA AGGATCTGTT GGAACAACCT ACTGTTACTC CAAACCGGAT
GGTAGACCAC CATCCACAGT GTCAGACCCA GTTACCAGAC TTGGACCCAC GCTTTCCAGG
CACTACACAT TTAAGGTAGG TGAGTGGCCC CATTCTCAAT CACATGGTCA CGCATGGATC
TGTCCGTTGC CAGGTGACAA ACTCAAGAAG ATGGGCAGTT TTCATGAGGT TGTCAAAGCC
CACCACCTGG TCAAGAACGG CTGGGATGTG GTTGTGCAGG TGAATCCCTC ATTTGCTCAC
TCCGGGCCGC TGTGTGTAGC AGCAGTGCCG GAGTACGAAC ACACACATGA GAAAGCACTC
AAGTGGTCTG AGCTTGAGGA ACCAGCTTAC ACATACCAAC AACTTTCAGT TTTTCCCCAC
CAGTTGCTAA ATTTGAGGAC AAATTCATCA GTGCATTTGG TGATGCCCTA CATTGGGCCA
GGCCAACCAA CAAATCTGAC TTTGCACAAC CCGTGGACCA TTGTTATTTT AATTTTGTCT
GAATTGACAG GACCTGGCCA AACTGTGCCT GTGACCATGT CGGTGGCTCC CATCGATGCA
ATGGTTAATG GGCCTCTTCC AAATCCAGAG

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

9. A substantially pure nucleotide sequence for VP3 being:

GCACCGATTA GAGTGGTGTC TGTGCCTGAA TCAGATTCTT TTATGTCTTC AGTACCTGAT
AATTCGACTC CACTATACCC CAAGGTTGTG GTCCCACCGC GCCAAGTTCC TGGCCGGTTT
ACAAATTTCA TTGATGTGGC AAAACAGACA TATTCATTTT GTTCCATTTC TGGAAAACCT
TATTTTGAGG TTACCAACAC CTCTGGGGAC GAGCCACTGT TTCAGATGGA TGTGTCGCTC
AGTGCGGCAG AGCTACATGG CACTTACGTA GCTAGTTTGT CATCATTTTT TGCACAGTAC
AGAGGCTCAC TTAATTTCAA CTTTATTTTC ACTGGTGCAG CAGCCACTAA GGCAAAGTTT
CTGGTTGCTT TTGTGCCTCC CCACAGTGCA GCGCCCAAAA CGCGCGATGA AGCAATGGCG
TGCATCCATG CCGTGTGGGA TGTTGGCTTG AACTCAGCTT TTTCTTTTAA TGTACCTTAT
CCCTCCCCTG CTGACTTCAT GGCCGTTTAT TCTGCGGAAC GGACGGTTGT GAATGTCTCT
GGATGGCTTC AAGTTTATGC ACTAACAGCT CTAACTTCAA CTGACATTGC CGTGAACAGT
AAAGGCCGTG TGCTGGTTGC TGTTTCCGCC GGCCCAGACT TCTCCCTTCG TCACCCGGCG
GACCTGCCCG ACAAGCAG

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

10. A substantially pure nucleotide sequence for VP4 being:

GGCGGAGGTA CATCCACTCC AACAACTGGC AACCAAAACA TGTCCGGAAA CAGTGGTTCA
ATTGTTCAAA ATTTTTACAT GCAACAGTAC CAGAATTCAA TTGACGCAGA CCTGGGAGAC
AATGTGATTA GCCCTGAAGG CCAGGGCAGC AACACTAGTA GTTCAACCTC ATCAAGCCAA
TCCTCTGGCT TGGGCGGGTG GTTCTCTAGT TTGCTGAACC TTGGAACAAA ACTACTGGCT

and functional equivalents of said nucleotide sequence including naturally occurring derivatives, variants and degeneracy equivalents.

11. Oligonucleotide primers derived from the nucleotide sequence of claim 1 being highly specific for ERhV1 or cross-reactive with other ERhV types.

12. An oligonucleotide primer according to claim 11 having the following nucleotide sequence:

VP1F 5' GTTGTGTTCAAGATTGCAGGC 3'

13. An oligonucleotide primer according to claim 11 having the following nucleotide sequence:

VP1R1 5' TTGCTCTCAACATCTCCAGC 3'

14. An oligonucleotide primer according to claim 11 having the following nucleotide sequence:

VP1R2 5' TAGCACCCTCCTTTATCATGCG 3'

15. Oligonucleotide probes derived from the nucleotide sequence of claim 1.

16. Diagnostic reagents, methods and kits characterised by the oligonucleotide primers and probes of claims 11 to 15.

17. Antigens comprising any one or a combination of the non-capsid proteins, being other than the individual VP1 to VP4 proteins, that are cleavage products of the polypeptide of claim 2.

18. Vaccines characterised by the incorporation of any one of a combination of virion proteins VP1 to VP4.

19. Vectors characterised by the incorporation of any one or a combination of virion proteins VP1 to VP4.

20. A diagnostic test for the detection of antibodies to ERhV1 in blood of horses and any other animal species characterised by the use of the antigens of claim 17.

21. A diagnostic test according to claim 20 being an enzyme linked immunosorbent assay.

22. A test to distinguish horses infected with ERhV1 in which said virus had replicated from horses which have been vaccinated with the vaccine of claim 18 comprising the steps of applying an antigen of claim 17 to a horse and testing for an immunoreaction thereto, wherein a positive immunoreaction would indicate that said horse had been infected with ERhV1 and a negative immunoreaction would indicate that said horse has not been infected with ERhV1.

23. Recombinant plasmids comprising nucleotide sequences and subsequences derived from the nucleotide sequence of claim 1.

24. A recombinant plasmid according to claim 22 comprising the P1-2A-3C region of the ERhV1 genome.

25. A host system characterised comprising the nucleotide sequence of claim 1 or part thereof.

26. A process for producing a protein product derived from ERhV1 comprising the steps of selecting out a gene of interest from the ERhV1 nucleotide sequence of claim 1 and expressing said protein product in a suitable host system.

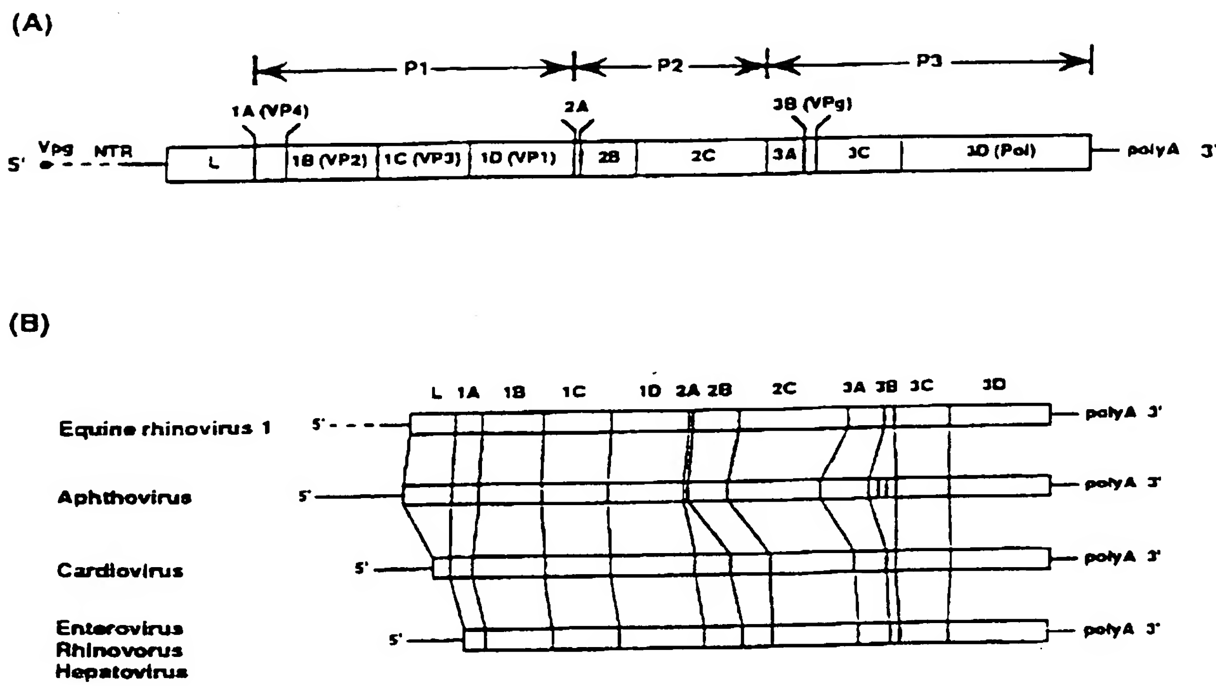

(A)
P1
P2
P3
1A (VP4)
2A
3B (VPg)
5' Vpg NTR
L
1B (VP2)
1C (VP3)
1D (VP1)
2B
2C
3A
3C
3D (Pol)
polyA 3'
(B)
L 1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
Equine rhinovirus 1
Aphthovirus
Cardiovirus
Enterovirus
Rhinovorus
Hepatovirus
5'
polyA 3'

FIGURE 1

```
CCGTCAAGCCCGTTGCCTGTATAGCCAGGTAACCGGACAGCGGCTTGCTGGATTTTCCCG GTGCCATTGCTCTGGATGGTGTCACCAAGCTGACAAATGCGGAGTGAACCTCACAAAGCG
 -430                 -410                  -390                 -370                  -350                 -330
ACACGCCTGTGGTAGCGCTGCCCAAAAGGGAGCGGAACTCCCCGCCGAGGCGGTCCTCTC TGGCCAAAAGCCCAGCGTTGATAGCGCCTTTTGGGATGCAGCAACCCCACCTGCCAGGTG
 -310                 -290                  -270                 -250                  -230                 -210
TGAAGTGGAGTGAGCGGATCTCCAATTTGGTCTGTTCTGAACTACACCATTTACTGCTGT GAAGAATGCCCTGGAGGCAAGCTGGTTACAGCCCTGACCAGGCCCTGCCCGTGACTCTCG
 -190                 -170                  -150                 -130                  -110                 -90
                                                                           L
                                                                           M  A  A  S  K  V  Y  R  V  C  E  Q  T  L  L
ACCGGCGCAGGGTCAAAAATTGTCTAAGCAGCAGCAGGAACGCGGGAGCGTTTCTTTTCC TTTTGTACTGACATGATGGCGGCGTCTAAGGTGTATAGAGTTTGCGAGCAGACTCTGCTG
 -70                  -50                   -30                  -10                   10                   30
A  G  A  V  R  M  M  D  K  F  L  Q  K  R  T  V  F  V  P  H   L  D  K  T  I  R  L  T  G  L  H  N  Y  D  N  T  C  W  L  N
GCAGGTGCCGTTCGCATGATGGACAAATTCTTGCAAAAGAGAACTGTTTTTGTCCCCCAT CTTGACAAAACAATTCGTTTGACTGGACTCCACAATTATGACAATACTTGCTGGTTGAAT
  50                   70                    90                   110                   130                  150
A  L  T  Q  L  T  Q  I  L  G  I  R  L  F  D  E  H  F  G  N   R  G  L  F  T  R  K  T  I  D  W  V  S  D  Q  T  G  I  K  D
GCCTTGACACAACTGACACAGATTCTTGGAATTCGGCTTTTTGATGAACACTTCGGCAAT AGAGGTCTGTTCACTCGGAAAACAATTGATTGGGTGAGTGACCAGACTGGTATAAAAGAT
  170                  190                   210                  230                   250                  270
L  K  S  G  A  P  P  L  V  V  V  Y  K  L  W  Q  H  G  H  L   D  V  G  T  M  E  K  P  R  S  I  T  L  W  S  G  P  K  V  C
CTAAAATCAGGAGCACCGCCACTCGTGGTGGTGTACAAACTGTGGCAACATGGACACTTG GATGTCGGTACGATGGAGAAACCCCGGTCGATTACTCTATGGTCTGGCCCCAAAGTGTGT
  290                  310                   330                  350                   370                  390
L  S  D  F  W  A  C  V  S  A  K  P  G  H  A  V  F  Y  L  L   T  S  E  G  W  I  C  V  D  D  K  K  I  Y  P  E  T  P  K  T
CTTTCTGATTTCTGGGCCTGTGTTTCGGCAAAACCGGGACATGCAGTATTCTACCTTCTC ACAAGCGAGGGTTGGATCTGTGTTGATGACAAGAAAATATACCCAGAAACACCCAAAACA
  410                  430                   450                  470                   490                  510
                                                                                             L ↓ VP4
E  D  V  L  V  F  A  P  Y  D  F  E  S  L  G  K  D  P  P  K   L  H  Q  R  Y  E  K  A  F  E  L  S  G  G  G  T  S  T  P  T
GAGGATGTACTTGTTTTTGCGCCCTATGACTTTGAGTCACTGGGCAAGGACCCACCAAAG CTACACCAGAGATATGAAAAAGCATTTGAGCTCAGTGGCGGAGGTACATCCACTCCAACA
  530                  550                   570                  590                   610                  630
T  G  N  Q  N  M  S  G  N  S  G  S  I  V  Q  N  F  Y  M  Q   Q  Y  Q  N  S  I  D  A  D  L  G  D  N  V  I  S  P  E  G  Q
ACTGGCAACCAAAACATGTCCGGAAACAGTGGTTCAATTGTTCAAAATTTTTACATGCAA CAGTACCAGAATTCAATTGACGCAGACCTGGGAGACAATGTGATTAGCCCTGAAGGCCAG
  650                  670                   690                  710                   730                  750
```

FIGURE 2a (1 of 7)

```
                                                                                                VP4 ↓ VP2
G  S  N  T  S  S  S  T  S  S  S  Q  S  S  G  L  G  G  W  F    S  S  L  L  N  L  G  T  K  L  L  A  D  K  K  T  E  E  T  T
GGCAGCAACACTAGTAGTTCAACCTCATCAAGCCAATCCTCTGGCTTGGGCGGGTGGTTC TCTAGTTTGCTGAACCTTGGAACAAAACTACTGGCTGACAAGAAGACAGAAGAGACTACA
  770                 790                 810                  830                 850                 870

N  I  E  D  R  I  E  T  T  V  V  G  V  T  I  I  N  S  Q  G    S  V  G  T  T  Y  C  Y  S  K  P  D  G  R  P  P  S  T  V  S
AACATTGAAGACAGAATTGAAACAACAGTGGTTGGAGTCACTATTATTAATTCACAAGGA TCTGTTGGAACAACCTACTGTTACTCCAAACCGGATGGTAGACCACCATCCACAGTGTCA
  890                 910                 930                  950                 970                 990

D  P  V  T  R  L  G  P  T  L  S  R  H  Y  T  F  K  V  G  E    W  P  H  S  Q  S  H  G  H  A  W  I  C  P  L  P  G  D  K  L
GACCCAGTTACCAGACTTGGACCCACGCTTTCCAGGCACTACACATTTAAGGTAGGTGAG TGGCCCCATTCTCAATCACATGGTCACGCATGGATCTGTCCGTTGCCAGGTGACAAACTC
  1010                1030                1050                 1070                1090                1110

K  K  M  G  S  F  H  E  V  V  K  A  H  H  L  V  K  N  G  W    D  V  V  V  Q  V  N  P  S  F  A  H  S  G  P  L  C  V  A  A
AAGAAGATGGGCAGTTTTCATGAGGTTGTCAAAGCCCACCACCTGGTCAAGAACGGCTGG GATGTGGTTGTGCAGGTGAATCCCTCATTTGCTCACTCCGGGCCGCTGTGTGTAGCAGCA
  1130                1150                1170                 1190                1210                1230

V  P  E  Y  E  H  T  H  E  K  A  L  K  W  S  E  L  E  E  P    A  Y  T  Y  Q  Q  L  S  V  F  P  H  Q  L  L  N  L  R  T  N
GTGCCGGAGTACGAACACACACATGAGAAAGCACTCAAGTGGTCTGAGCTTGAGGAACCA GCTTACACATACCAACAACTTTCAGTTTTTCCCCACCAGTTGCTAAATTTGAGGACAAAT
  1250                1270                1290                 1310                1330                1350

S  S  V  H  L  V  M  P  Y  I  G  P  G  Q  P  T  N  L  T  L    H  N  P  W  T  I  V  I  L  I  L  S  E  L  T  G  P  G  Q  T
TCATCAGTGCATTTGGTGATGCCCTACATTGGCCAGGCCAACCAACAAATCTGACTTTG CACAACCCGTGGACCATTGTTATTTTAATTTTGTCTGAATTGACAGGACCTGGCCAAACT
  1370                1390                1410                 1430                1450                1470
                                                               VP2 ↓ VP3
V  P  V  T  M  S  V  A  P  I  D  A  M  V  N  G  P  L  P  N    P  E  A  P  I  R  V  V  S  V  P  E  S  D  S  F  M  S  S  V
GTGCCTGTGACCATGTCGGTGGCTCCCATCGATGCAATGGTTAATGGGCCTCTTCCAAAT CCAGAGGCACCGATTAGAGTGGTGTCTGTGCCTGAATCAGATTCTTTTATGTCTTCAGTA
  1490                1510                1530                 1550                1570                1590

P  D  N  S  T  P  L  Y  P  K  V  V  V  P  P  R  Q  V  P  G    R  F  T  N  F  I  D  V  A  K  Q  T  Y  S  F  C  S  I  S  G
CCTGATAATTCGACTCCACTATACCCCAAGGTTGTGGTCCCACCGCGCCAAGTTCCTGGC CGGTTTACAAATTTCATTGATGTGGCAAAACAGACATATTCATTTTGTTCCATTTCTGGA
  1610                1630                1650                 1670                1690                1710
```

FIGURE 2a (2 of 7)

```
K  P  Y  F  E  V  T  N  T  S  G  D  E  P  L  F  Q  M  D  V   S  L  S  A  A  E  L  H  G  T  Y  V  A  S  L  S  S  F  F  A
AAACCTTATTTTGAGGTTACCAACACCTCTGGGGACGAGCCACTGTTTCAGATGGATGTG TCGCTCAGTGCGGCAGAGCTACATGGCACTTACGTAGCTAGTTTGTCATCATTTTTTGCA
  1730                 1750                 1770                  1790                 1810                 1830

Q  Y  R  G  S  L  N  F  N  F  I  F  T  G  A  A  A  T  K  A   K  F  L  V  A  F  V  P  P  H  S  A  A  P  K  T  R  D  E  A
CAGTACAGAGGCTCACTTAATTTCAACTTTATTTTCACTGGTGCAGCAGCCACTAAGGCA AAGTTTCTGGTTGCTTTTGTGCCTCCCCACAGTGCAGCGCCCAAAACGCGCGATGAAGCA
  1850                 1870                 1890                  1910                 1930                 1950

M  A  C  I  H  A  V  W  D  V  G  L  N  S  A  F  S  F  N  V   P  Y  P  S  P  A  D  F  M  A  V  Y  S  A  E  R  T  V  V  N
ATGGCGTGCATCCATGCCGTGTGGGATGTTGGCTTGAACTCAGCTTTTTCTTTTAATGTA CCTTATCCCTCCCCTGCTGACTTCATGGCCGTTTATTCTGCGGAACGGACGGTTGTGAAT
  1970                 1990                 2010                  2030                 2050                 2070

V  S  G  W  L  Q  V  Y  A  L  T  A  L  T  S  T  D  I  A  V   N  S  K  G  R  V  L  V  A  V  S  A  G  P  D  F  S  L  R  H
GTCTCTGGATGGCTTCAAGTTTATGCACTAACAGCTCTAACTTCAACTGACATTGCCGTG AACAGTAAAGGCCGTGTGCTGGTTGCTGTTTCCGCCGGCCCAGACTTCTCCCTTCGTCAC
  2090                 2110                 2130                  2150                 2170                 2190
                    VP3 ↓ VP1
P  A  D  L  P  D  K  Q  V  T  N  V  G  E  D  G  E  P  G  E   T  E  P  R  H  A  L  S  P  V  D  M  H  V  H  T  D  V  S  F
CCGGCGGACCTGCCCGACAAGCAGGTTACCAATGTGGGAGAGGATGGTGAACCCGGTGAG ACAGAGCCTCGTCATGCTTTGTCACCCGTGGACATGCACGTGCACACAGATGTCAGTTTC
  2210                 2230                 2250                  2270                 2290                 2310

L  L  D  R  F  F  D  V  E  T  L  E  L  S  N  L  T  G  S  P   A  T  H  V  L  D  P  F  G  S  T  A  Q  L  A  W  A  R  L  L
TTGCTTGACCGGTTCTTTGATGTTGAGACACTTGAGCTTTCAAATTTGACAGGTTCTCCT GCCACACATGTTCTGGATCCGTTTGGCTCGACTGCCCAACTGGCTTGGGCACGTCTGCTA
  2330                 2350                 2370                  2390                 2410                 2430

N  T  C  T  Y  F  F  S  D  L  E  L  S  I  Q  F  K  F  T  T   T  P  S  S  V  G  E  G  F  V  W  V  K  W  L  P  V  G  A  P
AACACTTGCACCTACTTCTTTTCTGATTTGGAATTGTCAATCCAGTTTAAATTTACCACC ACTCCGTCCTCTGTTGGAGAGGGCTTTGTGTGGGTGAAGTGGCTCCCTGTTGGAGCACCA
  2450                 2470                 2490                  2510                 2530                 2550
```

FIGURE 2a (3 of 7)

```
T  K  T  T  D  A  W  Q  L  E  G  G  G  N  S  V  R  I  Q  K   L  A  V  A  G  M  C  P  T  V  V  F  K  I  A  G  S  R  S  Q
ACCAAGACCACAGATGCTTGGCAGTTAGAAGGAGGTGGAAATTCAGTTAGAATTCAAAAA TTGGCCGTTGCAGGGATGTGCCCCACTGTTGTGTTCAAGATTGCAGGCTCCCGTTCACAA
  2570                2590                2610                2630                2650                2670

A  C  A  S  A  L  P  Y  T  S  M  W  R  V  V  P  V  F  Y  N   G  W  G  A  P  T  K  E  K  A  T  Y  N  W  L  P  G  A  H  F
GCCTGTGCTTCAGCGTTGCCATATACATCAATGTGGCGTGTTGTGCCAGTCTTTTACAAT GGCTGGGGTGCACCTACCAAAGAAAAGGCAACCTACAATTGGCTTCCTGGTGCACACTTT
  2690                2710                2730                2750                2770                2790

G  S  I  L  L  T  S  D  A  H  D  K  G  G  C  Y  L  R  Y  A   F  R  A  P  A  M  Y  C  P  R  P  I  P  P  A  F  T  R  P  A
GGTTCCATCTTGCTGACTTCTGATGCGCATGATAAAGGAGGGTGCTACTTGCGGTATGCT TTCCGCGCGCCAGCGATGTATTGCCCTCGACCCATTCCGCCGGCTTTTACGCGTCCAGCG
  2810                2830                2850                2870                2890                2910
                                        VP1 ↓ 2A                                                     2A ↓ 2B
D  K  T  R  H  K  F  P  T  N  I  N  K  Q  C  T  N  Y  S  L   L  K  L  A  G  D  V  E  S  N  P  G  P  T  I  F  S  K  A  S
GACAAAACCAGACATAAATTTCCCACTAACATCAACAAACAGTGTACTAATTACTCTCTC CTCAAATTGGCTGGAGATGTTGAGAGCAACCCTGGCCCCACTATTTTTTCCAAAGCATCA
  2930                2950                2970                2990                3010                3030

A  D  L  N  A  L  S  T  S  L  G  E  L  T  G  M  L  K  D  L   K  A  K  A  E  T  Y  S  P  F  Y  K  M  A  K  M  L  F  K  L
GCAGACCTGAATGCCTTGTCAACGTCGCTAGGTGAATTGACTGGCATGCTAAAAGATCTT AAAGCCAAGGCAGAAACTTATTCCCCGTTTTACAAAATGGCCAAAATGCTTTTCAAACTT
  3050                3070                3090                3110                3130                3150

A  T  L  A  V  A  A  M  R  T  K  D  P  V  V  V  V  M  L  I   A  D  F  G  L  E  V  F  D  T  G  F  F  F  S  Y  F  Q  E  K
GCAACACTAGCTGTGGCAGCTATGAGGACAAAGGACCCAGTAGTGGTGGTTATGTTGATT GCTGATTTCGGATTGGAGGTCTTTGACACTGGGTTTTTCTTTTCCTACTTTCAAGAGAAG
  3170                3190                3210                3230                3250                3270

L  Q  P  Y  M  K  T  I  P  G  K  I  S  D  L  V  T  D  A  A   T  A  A  A  Q  I  P  K  G  V  Y  S  F  V  S  S  F  F  E  T
TTGCAGCCTTATATGAAAACTATTCCTGGTAAGATTTCTGATTTGGTCACTGATGCGGCT ACGGCTGCCGCCCAAATTCCAAAGGGAGTGTATTCTTTTGTGTCGTCATTTTTCGAAACG
  3290                3310                3330                3350                3370                3390
                   2B ↓ 2C
P  E  G  V  V  E  K  Q  V  S  L  R  T  V  N  D  I  F  A  L   L  K  N  S  D  W  F  I  K  T  L  V  A  L  K  K  W  L  T  S
CCTGAAGGAGTGGTTGAGAAGCAGGTGTCTCTTCGGACAGTGAATGACATATTTGCTTTG CTTAAAAATTCTGATTGGTTCATAAAGACTCTTGTTGCCCTCAAGAAATGGCTGACATCC
  3410                3430                3450                3470                3490                3510
```

FIGURE 2a (4 of 7)

```
W  F  A  Q  E  Q  Q  A  D  D  A  L  Y  S  E  L  E  K  Y  P   L  Y  K  L  R  L  K  E  P  D  T  Q  E  E  A  R  Q  W  F  K
TGGTTTGCTCAAGAACAACAGGCAGATGATGCGCTCTATTCAGAATTGGAAAAATATCCC TTGTACAAGTTAAAATTGAAGGAACCTGATACTCAAGAGGAAGCGCGCCAGTGGTTTAAA
3530                3550                3570                3590                3610                3630

D  M  Q  Q  R  A  L  A  V  K  D  K  G  L  F  S  L  L  Q  I   P  L  V  N  L  P  Q  S  R  P  E  P  V  V  C  V  L  R  G  A
GACATGCAGCAGCGTGCTCTCGCTGTGAAGGACAAAGGTCTCTTTTCCCTCCTGCAAATT CCATTAGTTAACTTGCCCCAGAGCCGTCCAGAGCCCGTTGTATGCGTCCTTCGGGGCGCA
3650                3670                3690                3710                3730                3750

S  G  Q  G  K  S  Y  L  A  N  L  M  A  Q  A  I  S  L  L  L   V  G  K  Q  D  S  V  W  S  C  P  P  D  P  T  Y  F  D  G  Y
TCAGGGCAAGGCAAATCTTATTTGGCAAATCTGATGGCTCAAGCAATTTCGCTTCTCTTG GTTGGCAAGCAGGACAGTGTGTGGAGTTGTCCTCCTGACCCCACATATTTTGATGGCTAT
3770                3790                3810                3830                3850                3870

N  G  Q  A  V  V  I  M  D  A  L  G  Q  D  P  N  G  A  D  F   K  Y  F  C  Q  M  V  S  T  T  A  F  V  P  P  M  A  H  L  D
AACGGACAGGCTGTGGTGATTATGGATGCATTGGGCCAGGATCCGAATGGTGCTGACTTT AAATATTTTTGCCAGATGGTCTCTACAACAGCTTTTGTACCACCTATGGCCCATTTGGAT
3890                3910                3930                3950                3970                3990

D  K  G  I  P  F  T  S  P  V  V  I  C  T  T  N  L  H  S  S   F  T  P  I  T  V  S  C  P  E  A  L  K  R  R  F  R  F  D  V
GATAAAGGCATTCCATTTACTTCTCCTGTTGTTATTTGTACTACAAATTTGCATTCATCT TTTACCCCTATTACTGTTTCTTGTCCTGAAGCTCTTAAGAGGAGGTTTCGGTTTGATGTG
4010                4030                4050                4070                4090                4110

T  V  S  A  K  P  G  F  V  R  T  V  G  S  N  Q  L  L  N  L   P  L  A  L  K  P  A  G  L  P  P  H  P  I  F  E  N  D  M  P
ACGGTGTCCGCTAAACCGGGCTTTGTGCGCACTGTTGGTTCAAACCAGCTTTTGAATCTC CCACTTGCTCTTAAGCCAGCTGGTCTTCCCCCACACCCTATCTTTGAAAATGACATGCCC
4130                4150                4170                4190                4210                4230

I  I  N  G  Q  A  V  K  L  A  L  S  G  G  E  V  T  A  F  E   L  I  E  M  I  L  S  E  V  Q  N  R  Q  D  T  H  K  M  P  I
ATTATAAATGGGCAGGCTGTTAAATTGGCTCTTTCTGGTGGAGAAGTGACAGCTTTTGAG CTTATTGAGATGATACTGTCAGAAGTTCAAAACAGACAAGACACACACAAAATGCCCATT
4250                4270                4290                4310                4330                4350
  2C ↓ 3A
F  K  Q  S  W  S  D  L  F  R  K  C  T  T  D  E  E  Q  K  M   L  Q  F  L  I  D  N  K  D  S  E  I  L  R  A  F  V  S  E  R
TTTAAACAATCATGGTCTGATTTGTTCAGAAAGTGTACAACTGATGAGGAACAGAAAATG TTGCAGTTTTTAATTGACAATAAAGATTCAGAAATTCTCAGGGCGTTTGTTTCAGAACGC
4370                4390                4410                4430                4450                4470
```

FIGURE 2a (5 of 7)

```
S  I  L  L  H  E  E  Y  L  K  W  E  S  Y  M  T  R  R  A  K   F  H  R  L  A  A  D  F  A  M  F  L  S  I  L  T  S  L  I  V
TCCATTTTACTACATGAAGAGTATCTTAAATGGGAGTCATATATGACCAGGAGAGCCAAG TTTCACCGCCTGGCTGCTGATTTTGCTATGTTTCTATCCATTCTTACTTCACTGATTGTT
 4490                4510                4530                4550                4570                4590
                                                   3A ↓ 3B
I  F  C  L  V  Y  S  M  Y  Q  L  F  K  T  P  D  E  Q  S  A   Y  D  P  S  T  K  P  K  P  K  T  Q  E  V  K  T  L  K  I  R
ATTTTTTGTTTAGTTTATTCTATGTATCAACTTTTTAAGACCCCTGACGAGCAATCAGCT TATGATCCTTCAACTAAGCCAAAACCAAAGACCCAGGAAGTGAAAACACTGAAGATTAGG
 4610                4630                4650                4670                4690                4710
3B ↓ 3C
T  E  T  G  V  P  A  T  D  L  Q  Q  S  I  M  K  N  V  Q  P   I  E  L  Y  L  D  N  E  L  V  T  D  C  S  A  L  G  V  Y  D
ACTGAGACTGGTGTACCAGCAACTGACTTGCAACAATCCATCATGAAAAATGTTCAGCCA ATTGAGCTTTACCTTGACAATGAATTGGTTACTGACTGCTCTGCCTTGGGTGTTTATGAC
 4730                4750                4770                4790                4810                4830

N  S  Y  L  V  P  L  H  L  F  E  F  D  F  D  T  I  V  L  G   G  R  H  Y  K  K  A  E  C  E  K  V  E  F  E  L  E  V  H  G
AATTCATATTTGCTGCCCCTTCATTTGTTTGAATTTGATTTTGATACCATTGTGCTTGGT GGACGTCATTACAAGAAAGCTGAGTGTGAGAAGGTAGAGTTTGAGCTTGAAGTGAATGGA
 4850                4870                4890                4910                4930                4950

D  V  V  S  S  D  A  C  L  L  R  V  S  S  G  P  K  V  R  N   I  V  H  L  F  T  N  E  I  E  L  K  K  M  T  Q  V  T  G  I
GACGTGGTGTCATCAGATGCGTGTCTACTTCGAGTGTCATCGGGGCCTAAAGTTAGAAAT ATTGTTCATCTTTTTACAAATGAAATTGAATTGAAGAAAATGACCCAAGTGACAGGAATC
 4970                4990                5010                5030                5050                5070

M  N  S  P  H  Q  A  R  T  V  F  F  G  S  F  L  T  V  R  K   S  I  L  T  S  D  G  T  V  M  P  N  V  L  S  Y  A  A  Q  T
ATGAATTCACCACACCAGGCACGCACTGTGTTTTTTGGCAGTTTTTTGACAGTGAGGAAG TCCATCTTAACATCGGATGGACTGTAATGCCCAATGTTTTGTCCTATGCCGCTCAGACC
 5090                5110                5130                5150                5170                5190

S  R  G  Y  C  G  A  A  I  V  A  G  S  P  A  R  I  I  G  I   H  S  A  G  T  G  S  V  A  F  C  S  L  V  S  R  D  A  L  E
TCGCGTGGGTATTGTGGCGCTGCAATTGTTGCTGGCTCACCTGCCCGCATAATTGGTATC CATTCAGCTGGCACTGGATCTGTTGCATTTTGCTCCCTGGTGTCCAGAGACGCGCTGGAG
 5210                5230                5250                5270                5290                5310
          3C ↓ 3D
Q  L  W  P  Q  K  Q  G  N  V  S  R  L  D  D  D  V  R  V  S   V  P  R  R  S  K  L  V  K  S  L  A  Y  P  I  F  K  P  D  Y
CAACTCTGGCCCCAGAAACAGGGCAACGTTAGTCGCCTTGATGACGATGTGAGGGTGTCT GTTCCGCGCCGCTCCAAATTGGTGAAATCATTGGCTTACCCCATTTTCAAACCTGACTAT
 5330                5350                5370                5390                5410                5430
```

# FIGURE 2a (6 of 7)

```
G P A P L S Q F D K R L S D G V K L D E  V V F A K H T G D K E I S A Q D Q K W L
GGCCCAGCGCCACTCTCTCAATTTGACAAGCGCCTGTCAGACGGCGTGAAGCTGGATGAA GTGGTTTTTGCTAAACATACTGGAGACAAGGAGATTTCCGCACAGGACCAGAAATGGCTC
 5450                5470                5490                5510                5530                5550

L R A A H V Y A Q K V F S R I G F D N Q  A L T E K E A I C G I P G L D K M E Q D
TTGCGTGCGGCGCATGTATACGCCCAGAAGGTTTTCTCCCGGATTGGATTTGACAACCAG GCTTTGACTGAAAAAGAGGCCATTTGTGGCATTCCTGGCCTTGACAAGATGGAGCAGGAC
 5570                5590                5610                5630                5650                5670

T A P G L P Y A Q Q N K R R K D I C D F  E E G R L K G A E L Q K D R F M A G D Y
ACCGCTCCCGGGCTGCCCTATGCTCAGCAAAATAAGAGAAGGAAAGACATCTGTGATTTT GAAGAGGGCCGGCTGAAGGGCGCCGAACTCCAAAAGGACAGATTTATGGCTGGTGACTAC
 5690                5710                5730                5750                5770                5790

S N L V Y Q S F L K D E I R P L E K V R  A G K T R L I D V P P M P H V V V G R Q
TCTAATTTGGTCTATCAATCATTTTTGAAAGATGAGATCCGCCCACTTGAGAAAGTTAGG GCTGGAAAGACCCGCCTGATTGACGTGCCGCCGATGCCCCATGTGGTGGTTGGTAGGCAG
 5810                5830                5850                5870                5890                5910

L L G R F V A K F H E A N G F D I G S A  I G C D P D V D W T R F G L E L E R F R
CTCTTGGGCCGGTTTGTGGCAAAATTTCATGAAGCAAATGGATTTGACATTGGCTCAGCC ATTGGATGTGACCCAGATGTGGACTGGACTCGGTTTGGCCTCGAGTTGGAGCGTTTCAGG
 5930                5950                5970                5990                6010                6030

Y V Y A C D Y S R F D A N H A A D A M R  V V L N Y F F S E D H G F D P G V P A F
TATGTATATGCCTGTGACTACTCACGGTTCGATGCCAACCATGCAGCTGATGCAATGAGA GTTGTGCTTAACTACTTTTTCTCTGAGGACCACGGTTTCGACCCTGGTGTGCCTGCTTTT
 6050                6070                6090                6110                6130                6150

I E S L V D S V H A Y E E K R Y N I Y G  G L P S G C S C T S I L N T I L N N V Y
ATTGAGTCACTGGTTGATTCAGTGCATGCCTATGAAGAGAAAAGGTATAACATCTACGGT GGCTTGCCATCCGGGTGTTCCTGCACATCAATTTTGAATACCATCTTGAACAATGTTTAC
 6170                6190                6210                6230                6250                6270

I L A A M M K A Y E N F E P D D I Q V I  C Y G D D C L I A S D F E I D F Q Q L V
ATTCTTGCAGCTATGATGAAGGCTTATGAGAATTTTGAGCCAGATGACATTCAGGTCATT TGCTATGGGGACGACTGCCTCATTGCTTCTGATTTTGAAATTGATTTCCAACAACTGGTG
 6290                6310                6330                6350                6370                6390

P V F S S F G Q V I T T A D K T D F F K  L T T L S E V T F L K R A F V L T A F Y
CCTGTCTTTTCTAGTTTTGGACAGGTAATAACTACAGCTGACAAGACTGATTTTTTTAAA CTGACAACGCTTTCGGAGGTGACCTTCCTTAAGCGCGCTTTTGTTCTGACGGCCTTTTAC
 6410                6430                6450                6470                6490                6510

K P V M D V K T L E A I L S F V R P G T  Q A E K L L S V A Q L A G H C E P E Q Y
AAGCCAGTGATGGATGTGAAGACCCTTGAAGCAATCTTAAGCTTTGTTCGCCCAGGCACA CAGGCTGAAAAGCTCCTGTCCGTGGCGCAGTTGGCAGGCCACTGCGAACCGGAGCAGTAT
 6530                6550                6570                6590                6610                6630

                                                                               3D
E R L F E P F A G M Y F V P T W R L A P  A V V D E A W M L N S F
GAGCGCCTGTTTGAGCCCTTTGCTGGGATGTATTTCGTCCCTACTTGGCGACTTGCGCCT GCAGTGGTTGATGAAGCTTGGATGCTAAATTCTTTTTGACTTTGTTTTTCTTTGTTTTCT
 6650                6670                6690                6710                6730                6750

TTTAGGCTTTTAAGGTGTTAAGTTTAAAGGTTAAGAGTTTTTAGAAGTTAAGATAGAGTT TAGTTTTTAGTTTTGAGC-poly(A)
 6770                6790                6810                6830
```

FIGURE 2a (7 of 7)

# FIGURE 2b

-790 -770 -750
TAAGTAAAACGCTGTAACTGCATGATTTGCGCCTGTAGCGCCAGTAAAACGCAGAAACCA
-730 -710 -690
CAAGCAAAAACCTGTAGCGTCAGTAAAACGCGCACATTCACATACAGAGCTTCCCGGCTT
-670 -650 -630
TAAGGGTTACTGCTCGTAATGAGAGCACATGACAACTTGTCGAGATTACGGCAACTGTCA
-610 -590 -570
CGGGAGAGAGGAGCCCGTTTTCGGGCACTTGTCTCCTAAACAATGTTGGCGCGCATTTGC
-550 -530 -510
GCGCCCCCCCCCTTTTCAGCCCCCTGTCATTGACTGGTCGAAGCGTTCGCAATAAGACT
-490 -470 -450
GGTCGTCACTTGGCTGTTCTATCGTTTCAGGCTTTAGCGCGCCCTTGCGCGGCGGGCCGT
-430 -410 -390
CAAGCCCGTGCGCTGTATAGCGCCAGGTAACCGGACAGCGGCGTGCTGGATTTTCCCGGT
-370 -350 -330
GCCATTGCTCTGGATGGTGTCACCAAGCTGACAAATGCGGAGTGAACCTCACAAAGCGAC
-310 -290 -270
ACGCCTGTGGTAGCGCTGCCCAAAAGGGAGCGGAACTCCCCGCCGAGGCGGTCCTCTCTG
-250 -230 -210
GCCAAAAGCCCAGCGTTGATAGCGCCTTTTGGGATGCAGGAACCCCACCTGCCAGGTGTG
-190 -170 -150
AAGTGGAGTGAGCGGATCTCCAATTTGGTCTGTTCTGAACTACACCATTTACTGCTGTGA
-130 -110 -90
AGAATGCCCTGGAGGCAAGCTGGTTACAGCCCTGACCAGGCCCTGCCCGTGACTCTCGAC
-70 -50 -30
CGGCGCAGGGTCAAAAATTGTCTAAGCAGCAGCAGGAACGCGGGAGCGTTTCTTTTCCTT
-10 10 30
TTGTACTGACATGATGGCGGCGTCTAAGGTGTATAGAGTTTGCGAGCAGACTCTGCTGGC
M A A S K V Y R V C E Q T L L A
50 70
**AGGTGCCGTTCGCATGATGGACAAA**
G A V R M M D K

```
FMDVO1K     1 MNTTDCFIALVQAIREIKALFLSRTTGKMELTLYNGEKKTFYSRPNNHDN-CWLNAILQL  59
              *.....:     *.:        ::.:  .*..: : : :*.. ..   :*.** *****: **
ERhV1       1 MAASKVYRVCEQTLLAGAVRMMDKFLQKRTVFVPHLDKTIRLTGLHNYDNTCWLNALTQL  60

FMDVO1K    60 FRYVEEPFFDWVYSSPENLTLEAIKQLEDLTGL-ELHEGGPPALVIWNIKHLLHTGIGTA  118
               .:: .:**  ::..: :* ..*. :.* **: :*..*:** :*:::.:.:  * ::**
ERhV1      61 TQILGIRLFDEHFGNRGLFTRKTIDWVSDQTGIKDLKSGAPPLVVVYKLWQHGHLDVGTM  120

FMDVO1K   119 SRPSEVCMVDGTDMCLADFHAGIFLKGQEHAVFACVTSNGWYAIDDEDFYPWTPDPSDVL  178
               :*..:.: .*..:**.** *.:  *  .:****  :**:**..:**..:** **...***
ERhV1     121 EKPRSITLWSGPKVCLSDFWACVSAK-PGHAVFYLLTSEGWICVDDKKIYPETPKTEDVL  179

                                    L ↓ VP4
FMDVO1K   179 VFVPYDQEPLNGEWKAKVQR-----KLKGAGQSSPATGSQNQSGNTGSIINNYYMQQYQN  233
              **.*** *.*. :      **      .*.*:* *.*.**.** ***.***:.*:*******
ERhV1     180 VFAPYDFESLGKDPPKLHQRYEKAFELSGGGTSTPTTGNQNMSGNSGSIVQNFYMQQYQN  239
                                    L ↑ VP4
                                                             VP4 ↓ VP2
FMDVO1K   234 SMDTQLGDNAISGGSNEGSTDTTSTHTTNTQNNDWFSKLASSAFSGLFGALLADKKTEET 293
              *:*.:****.**.::.::.*.....     ... .:***.*      *    **********
ERhV1     240 SIDADLGDNVISPEGQGSNTSSSTSSSQSSGLGGWFSSLL-----NLGTKLLADKKTEET 294
                                                             VP4 ↑ VP2

FMDVO1K   294 TLLEDRILTTRNGHTTSTTQSSVGVTYGYATAEDFVSGPNTSGLETRVVQAERFFKTHLF 353
              * :**** **   * *. ..*:***.**.*..::: ..:. ...:.  ..  .* :. .:
ERhV1     295 TNIEDRIETTVVGVTIINSQGSVGTTYCYSKPDGRPPSTVSDPVTRLGPTLSRHYTFKVG 354

FMDVO1K   354 DWVTSDSFGRCHLLELPTDHKGVYGSLTD---SYAYMRNGWDVEVTAVGNQFNGGCLLVA 410
              :*. *:* *:. : .**.*.      **: :    .. .::***** *    ..   ::*.* **
ERhV1     355 EWPHSQSHGHAWICPLPGDKLKKMGSFHEVVKAHHLVKNGWDVVVQVNPSFAHSGPLCVA 414
```

FIGURE 3 (1 of 6)

```
FMDVO1K   411 MVPELYSIQKREL------------YQLTLFPHQFINPRTNMTAHITVPFVGVNRYDQYK 458
                ***.   .:.:.*            **.:****;:* ***  ..*:.:*:*... ....
ERhV1     415 AVPEYEHTHEKALKWSELEEPAYTYQQLSVFPHQLLNLRTNSSVHLVMPYIGPGQPTNLT 474

                                                  VP2 ↓ VP3
FMDVO1K   459 VHKPWTLVVMVVAPLTVNTEGAPQIKVYANIAPTNVHVAGEFPSKEGIFPVACSDGYGGL 518
              :*.***:*:::::..**. .:..*    *   .:**.:. *.*.:*..*: :.*.: .: :::
ERhV1     475 LHNPWTIVILILSELTGPGQTVP---VTMSVAPIDAMVNGPLPNPEAPIRVVSVPESDSF 531
                                                  VP2 ↑ VP3

FMDVO1K   519 VTTDPKTADPVYGKVFNPPRNQLPGRFTNLLDVAEACPTFLRFEGGVPYVTTKTDSDRVL 578
              :.. *....*:*.**. *** *:******:;***... .* .:.*   : .*.*.:* .*
ERhV1     532 MSSVPDNSTPLYPKVVVPPR-QVPGRFTNFIDVAKQTYSFCSISGKPYFEVTNTSGDEPL 590

FMDVO1K   579 AQFDMSLAAKQMSNTFLAGLAQYYTQYSGTINLHFMFTGPTDAKARYMVAYAPPGMEPPK 638
                *:*:**.* :: .*::*:*. ::.**.*.:*::*:***:...**:::**:.**  .:**
ERhV1     591 FQMDVSLSAAELHGTYVASLSSFFAQYRGSLNFNFIFTGAAATKAKFLVAFVPPHSAAPK 650

FMDVO1K   639 TPEAAAHCIHAEWDTGLNSKFTFSIPYLSAADYAYTASGVAETTNVQGWVCLFQITHGKA 698
              *.:.*  **** **.**** *.*.:** *:**:    *:   ...** **: :::*  ..
ERhV1     651 TRDEAMACIHAVWDVGLNSAFSFNVPYPSPADFMAVYSAERTVVNVSGWLQVYALTALTS 710

                                        VP3 ↓ VP1
FMDVO1K   699 DGDA------LVVLASAGKDFELRLPVD-ARAETTSAGESADPVTTTVENYGGETQIQRR 751
               : *       ::* .***.**.** *.* :  :.*..**.::*..*.. :  : ..:::.:
ERhV1     711 TDIAVNSKGRVLVAVSAGPDFSLRHPADLPDKQVTNVGEDGEPGETEPRH--ALSPVDMH 768
                                        VP3 ↑ VP1

FMDVO1K   752 QHTDVSFIMDRFVKV--------TPQNQINILDLMQIPSHTLVGALLRASTYYFSDLEIA 803
                ******:;***..*        *.  ..::** :. ..:  : **..:**:*****:.
ERhV1     769 VHTDVSFLLDRFFDVETLELSNLTGSPATHVLDPFGSTAQLAWARLLNTCTYFFSDLELS 828
```

FIGURE 3 (2 of 6)

```
FMDVO1K   804 VKHEGDLT------------WVPNGAPEKALDNTTNPTAYHKAPLTRLALPYTAPHRVL-  850
              :.   . *             *:* ***.*. *.     ..: :...: :**::  .*  *:
ERhV1     829 IQFKFTTTPSSVGEGFVWVKWLPVGAPTKTTDAWQLEGGGNSVRIQKLAVAGMCPTVVFK 888

FMDVO1K   851 -------ATVYNGECRYNRNAVPNLRGDLQVLAQKVAR-------TLPTSFNYGAIKATR 896
                     *..    .:    :..** :  .::.. ... *          :.. :   :. ...
ERhV1     889 IAGSRSQACASALPYTSMWRVVPVFYNGWGAPTKEKATYNWLPGAHFGSILLTSDAHDKG 948

                                                 VP1 ↓ 2A           2A ↓ 2B
FMDVO1K   897 VTELLYRMKRAETYCPRPL-LAIHPTEARHKQKIVAPV-KQTLNFDLLKLAGDVESNPGP 954
                * *  :: :. *****:  *:  ....: ::*:.. : **.  *:.***************
ERhV1     949 GCYLRYAFRAPAMYCPRPIPPAFTRPADKTRHKFPTNINKQCTNYSLLKLAGDVESNPGP 1008
                                                 VP1 ↑ 2A           2A ↑ 2B

FMDVO1K   955 FFFSDVRSNFSKLVETINQMQEDMSTKHGPDFNRLVSAFEELAIGVKAIRTGLDEAKPWY 1014
               :**                       * ::*:* *  ..::**.   :*.::.  :. .*:*
ERhV1    1009 TIFS---------------------KASADLNALSTSLGELTGMLKDLKAKAETYSPFY 1046

FMDVO1K  1015 KLIKLLSRLSCMAAVAARSKDPVLVAIMLADTGLEILDSTFVVKKISDSLSSLFHVPAPV 1074
              *: *:* :*..:*..* *.****:*.:::** ***::*..*...  : :.* ..:.. :.
ERhV1    1047 KMAKMLFKLATLAVAAMRTKDPVVVVMLIADFGLEVFDTGFFFSYFQEKLQPYMKTIPGK 1106

                                         2B ↓ 2C
FMDVO1K  1075 FS---FGAPVLLAGLVKVASSFFRSTPEDLE-RAEKQLKARDINDIFAILKNGEWLVKLI 1130
              :*    :*:.  *.:.*..  **..*  *. *  .***:. *.:*****:***::*::* :
ERhV1    1107 ISDLVTDAATAAAQIPKGVYSFVSSFFETPEGVVEKQVSLRTVNDIFALLKNSDWFIKTL 1166
                                         2B ↑ 2C

FMDVO1K  1131 LAIRDWIKAWIASEEKF-VTMTDLVPGILEKQRDLNDPSKYKEAKEWLDNARQACLKSGN 1189
              :*::.*:..*:* *:.   .:   :.        *.:*.   **::*:.: .* .*    :
ERhV1    1167 VALKKWLTSWFAQEQQADDALYSELEKYPLYKLKLKEPDTQEEARQWFKDMQQRALAVKD 1226
```

FIGURE 3 (3 of 6)

```
FMDVO1K  1190 VHIANLCKVVAPAPSKSRPEPVVVCLRGKSGQGKSFLANVLAQAISTHFTGRIDSVWYCP 1249
                : .*  .:. .. ..*******..*** ******:***::*****  :.*: **** **
ERhV1    1227 KGLFSLLQIPLVNLPQSRPEPVVCVLRGASGQGKSYLANLMAQAISLLLVGKQDSVWSCP 1286

FMDVO1K  1250 PDPDHFDGYNQQTVVVMDDLGQNPDGKDFKYFAQMVSTTGFIPPMASLEDKGKPFNSKVI 1309
              ***..*****.*.**:**.***:*:* *****.******:*:**** *:*** **.*.*:
ERhV1    1287 PDPTYFDGYNGQAVVIMDALGQDPNGADFKYFCQMVSTTAFVPPMAHLDDKGIPFTSPVV 1346

FMDVO1K  1310 IATTNLYSGFTPRTMVCPDALNRRFHFDIDVSAKDGY----KINSKLDIIKALEDTHANP 1365
              *.****.*:*** *: **:**.***:**:.****.*:      *  *::  **...   *
ERhV1    1347 ICTTNLHSSFTPITVSCPEALKRRFRFDVTVSAKPGFVRTVGSNQLLNLPLALKPAGLPP 1406

                                                                2C ↓ 3A
FMDVO1K  1366 VAMFQYDCALLNGMAVEMKRMQQDMFKPQPPLQNVYQLVQEVIDRVELHEKVSSHPIFKQ 1425
                ::*: * :::** **.:  ..::          ::*:: ::. *: ::..  *****
ERhV1    1407 HPIFENDMPIINGQAVKLALSGGEV--------TAFELIEMILSEVQNRQDTHKMPIFKQ 1458
                                                                2C ↑ 3A

FMDVO1K  1426 ISIPSQKSVLYFLIEKGQHEAAIEFFEGMVHDSIKEELRPLIQQTSFVKRAFKRLKENFE 1485
                          *: *.  :.. .::: ::.:. .* **::: : *:: :.   .*.:
ERhV1    1459 SWSD--------LFRKCTTDEEQKMLQFLIDNKDSEILRAFVSERSILLHEEYLKWESYM 1510

FMDVO1K  1486 IVALCLTLLANIVIMIRETRKRQKMVDDAVNEYIEKANITTDDKTLDEAEKSPLETSGAS 1545
                   :  **.  *: .. .. ::  *   .:  ....::...*...*.. .* :..
ERhV1    1511 TRRAKFHRLAADFAMFLSILTSLIVIFCLVYSMYQLFKTPDEQSAYDPSTKPKPKTQEVK 1570
                                                    3A ↑ 3B

                                       3A ↓ 3B                  3A ↓ 3B
FMDVO1K  1546 TVGFRERTLPGQKACDDVNSEPAQPVEEQPQAEGPYAGPLERQKPLKVRAKLPQQEGPYA 1605
              *: :*
ERhV1    1571 TLKIR------------------------------------------------------- 1575
```

FIGURE 3 (4 of 6)

```
                         3A ↓ 3B                                3B ↓ 3C
FMDVO1K  1606  GPMERQKPLKVKAKAPVVKEGPYEGPVKKPVALKVKAKNLIVTESGAPPTDLQKMVMGNT 1665
                                                                **.*.*:****. :* *.
ERhV1    1576  ------------------------------------------TETGVPATDLQQSIMKNV 1593
                                                                3B ↑ 3C

FMDVO1K  1666  KPVELILDGKTVAICCATGVFGTAYLVPRHLFAEKYDKIMVDGRAMTDSDYRVFEFEIKV 1725
                *:**.**.. *. *:* **:::..**** ***. .:*.*:::**  ...::  .***:.*
ERhV1    1594  QPIELYLDNELVTDCSALGVYDNSYLVPLHLFEFDFDTIVLGGRHYKKAECEKVEFELEV 1653

FMDVO1K  1726  KGQDMLSDAALMVLHRGNRVRDITKHFRDTARMKKGTPVVGVINNADVGRLIFSGEALTY 1785
                *: : ***.*: : .* :**:*.. * :.  :** *.*.*::*.:. :* :* *. **
ERhV1    1654  NGDVVSSDACLLRVSSGPKVRNIVHLFTNEIELKKMTQVTGIMNSPHQARTVFFGSFLTV 1713

FMDVO1K  1786  KDIVVCMDGDTMPGLFAYRAATKAGYCGGAVLAKDGADTFIVGTHSAGGNGVGYCSCVSR 1845
               :. ::. **..**.::.* *.*. ****:*::*  *...  *:*.****..:*::** ***
ERhV1    1714  RKSILTSDGTVMPNVLSYAAQTSRGYCGAAIVA--GSPARIIGIHSAGTGSVAFCSLVSR 1771

                      3C ↓ 3D
FMDVO1K  1846  SMLLKMKAHIDPEPHHEGLIVDTRDVEERVHVMRKTKLAPTVAHGVFNPEFGPAALSNKD 1905
                * .:.:: :.:        *.  * : ** * *:.**...:*..:*.*::***:**. *
ERhV1    1772  DALEQLWPQKQGN-------VSRLDDDVRVSVPRRSKLVKSLAYPIFKPDYGPAPLSQFD 1824
                  3C ↑ 3D

FMDVO1K  1906  PRLNEGVVLDEVIFSKHKGDTKMSEEDKALFRRCAADYASRLHSVLGTANAPLSIYEAIK 1965
                **.:** ****:*.**.**..:*.:*. :: *.*  ** :: * :* .*.:*.  ***
ERhV1    1825  KRLSDGVKLDEVVFAKHTGDKEISAQDQKWLLRAAHVYAQKVFSRIGFDNQALTEKEAIC 1884

FMDVO1K  1966  GVDGLDAMEPDTAPGLPWALQGKRRGALIDFENGTVGPEVEAALKLMEKREYKFVCQTFL 2025
                *:.*** **.*******:* *.*** .:.***:* : ..      ::*.    .:*:*.**
ERhV1    1885  GIPGLDKMEQDTAPGLPYAQQNKRRKDICDFEEGRLKGAELQKDRFMAGDYSNLVYQSFL 1944
```

FIGURE 3 (5 of 6)

```
FMDVO1K  2026 KDEIRPLEKVRAGKTRIVDVLPVEHILYTRMMIGRFCAQMHSNNGPQIGSAVGCNPDVDW 2085
              ****************::** *:.*:: .* ::***.*.:*..** :****:**:*****
ERhV1    1945 KDEIRPLEKVRAGKTRLIDVPPMPHVVVGRQLLGRFVAKFHEANGFDIGSAIGCDPDVDW 2004

FMDVO1K  2086 QRFGTHFAQYRNVWDVDYSAFDANHCSDAMNIMFEEVFRTEFGFHPNAEWILKTLVNTEH 2145
               ***  .:..:* *:..*** *****..***.:::: .*..: **.*... ::..**:. *
ERhV1    2005 TRFGLELERFRYVYACDYSRFDANHAADAMRVVLNYFFSEDHGFDPGVPAFIESLVDSVH 2064

FMDVO1K  2146 AYENKRITVGGGMPSGCSATSIINTILNNIYVLYALRRHYEGVELDTYTMISYGDDIVVA 2205
              ***:**..: **:*****.***:******:*:* *:.: **..* *.. :*:****.::*
ERhV1    2065 AYEEKRYNIYGGLPSGCSCTSILNTILNNVYILAAMMKAYENFEPDDIQVICYGDDCLIA 2124

FMDVO1K  2206 SDYDLDFEALKPHFKSLGQTITPADKSDKGFVLGHSITDVTFLKRHFHMDYGTGFYKPVM 2265
              **:::**:.* * *.*:**.**.***.*  *.   .:.:****** * :    *:******
ERhV1    2125 SDFEIDFQQLVPVFSSFGQVITTADKTD--FFKLTTLSEVTFLKRAFVL---TAFYKPVM 2179

FMDVO1K  2266 ASKTLEAILSFARRGTIQEKLISVAGLAVHSGPDEYRRLFEPFQGLFEIPSYR        2318
                 *********.*.** .***:***.**.*::*::* ******.*:: :*.:*
ErhV1    2180 DVKTLEAILSFVRPGTQAEKLLSVAQLAGHCEPEQYERLFEPFAGMYFVPTWR        2232
```

FIGURE 3 (6 of 6)

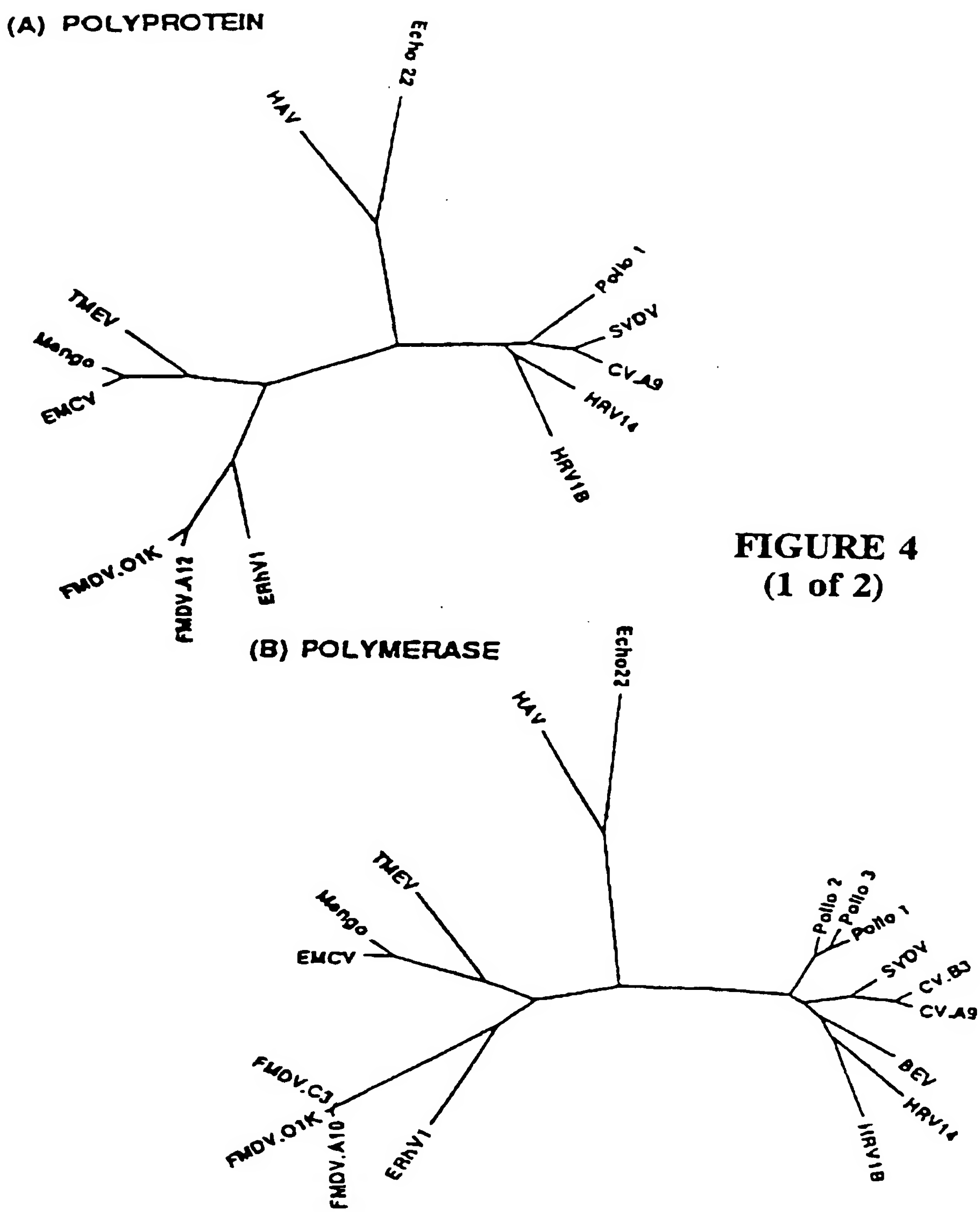



FIGURE 4
(1 of 2)

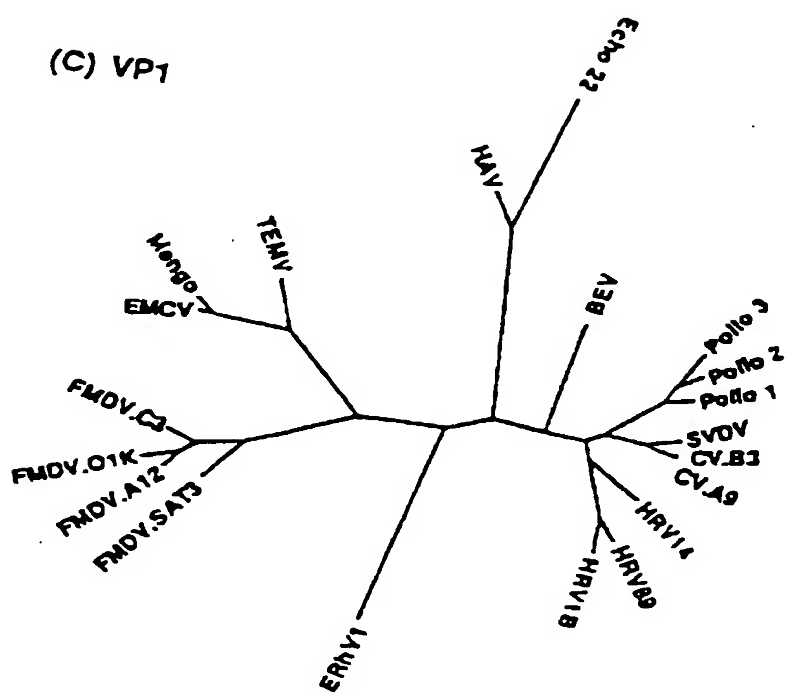
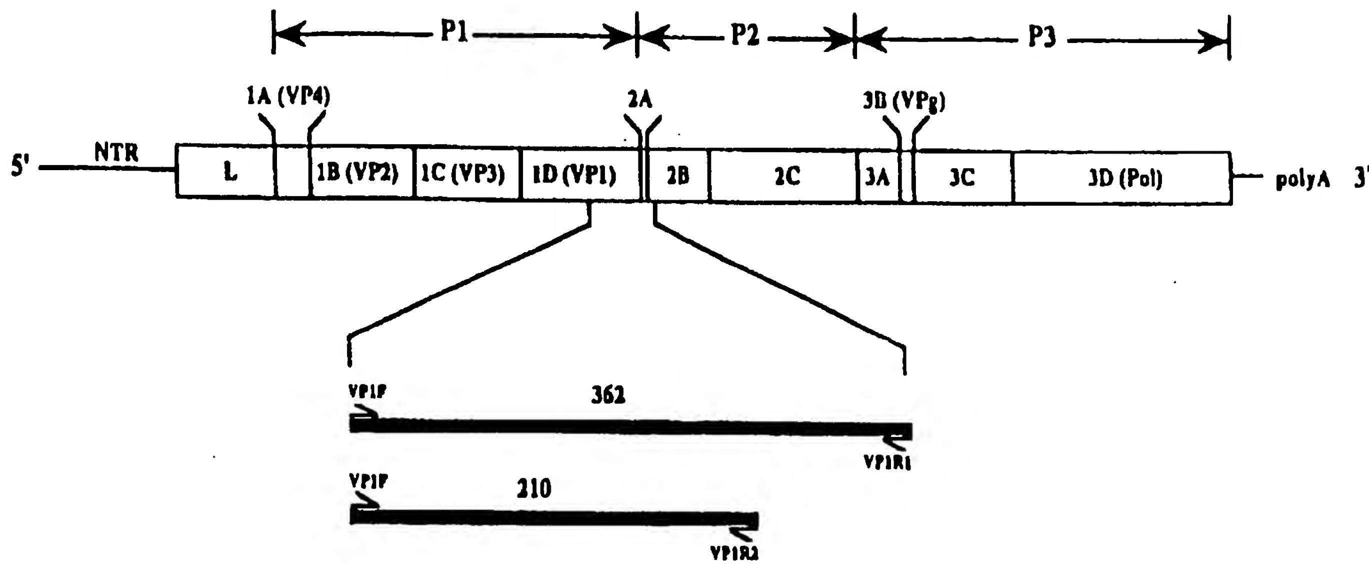



**FIGURE 4**
**(2 of 2)**

(A)




(B)

| | | | |
|---|---|---|---|
| VP1F | 5' | GTTGTGTTCAAGATTGCAGGC | 3' |
| VP1R1 | 5' | TTGCTCTCAACATCTCCAGC | 3' |
| VP1R2 | 5' | TAGCACCCTCCTTTATCATGCG | 3' |

FIGURE 5

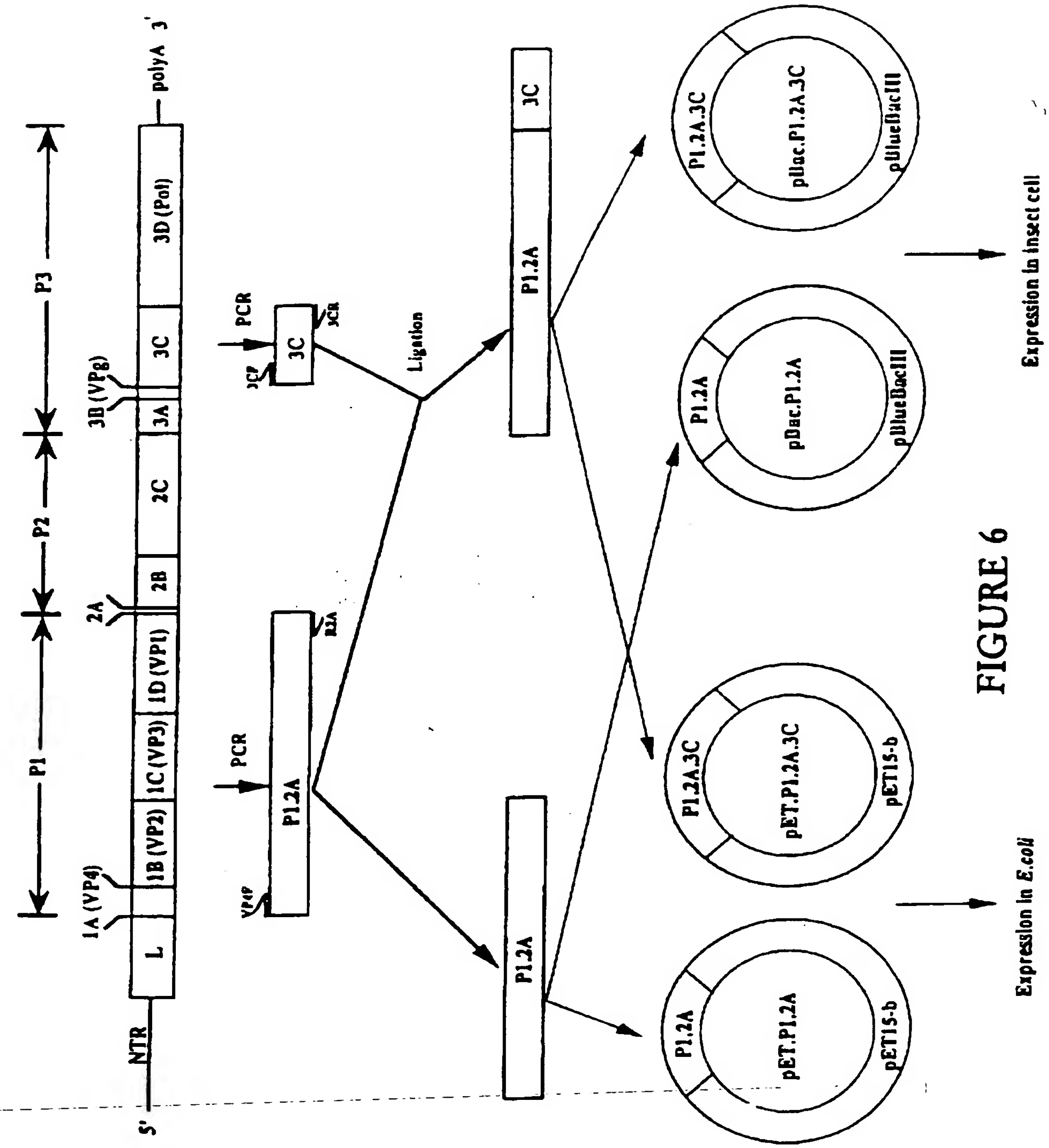

5'
NTR
L
1A (VP4)
1B (VP2)
1C (VP3)
1D (VP1)
2A
2B
2C
3A
3B (VPg)
3C
3D (Pol)
polyA 3'
P1
P2
P3
PCR
P1.2A
VP4P
R2A
3CP
3CR
Ligation
P1.2A
3C
P1.2A
pET.P1.2A
pET15-b
P1.2A.3C
pET.P1.2A.3C
pET15-b
P1.2A
pBac.P1.2A
pBlueBacIII
P1.2A.3C
pBac.P1.2A.3C
pBlueBacIII
Expression in *E.coli*
Expression in insect cell

FIGURE 6

## INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 96/00815

| A. | CLASSIFICATION OF SUBJECT MATTER |
|---|---|
| Int Cl[6]: | C12N 15/41; C07K 14/095; A61K 39/125; G01N 33/53, 33/569; C12Q 1/68 |

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
WPAT; JAPIO; CHEMICAL ABSTRACTS : KEYWORDS AS BELOW

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPM; STN: GENBANK; SWISS-PROT

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT, JAPIO, CA, USPM-EQUINE(N)RHINOVIR: OR ERH OR ERHV GENBANK, SWISSPROT-FULL NUCLEOTIDE AND AMINO ACID SEQUENCES

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| PX | LI F. ET AL. (1996) "EQUINE RHINOVIRUS 1 IS MORE CLOSELY RELATED TO FOOT-AND-MOUTH DISEASE VIRUS THAN TO OTHER PICORNAVIRUSES" Proc. Natl. Acad. Sc. USA Vol. 93 pp 990-995. See entire document | 1-26 |
| PX | WUTZ G. ET AL (1996) "EQUINE RHINOVIRUS SEROTYPES 1 AND 2 : RELATIONSHIP TO EACH OTHER AND TO APTHOVIRUSES AND CARDIOVIRUSES" Journal of General Virology Vol 77 pp 1719-1730 See entire document | 1-26 |

[X] Further documents are listed in the continuation of Box C

[ ] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 30 January 1997 | 18 FEB 1997 |
| Name and mailing address of the ISA/AU<br>AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION<br>PO BOX 200<br>WODEN ACT 2606<br>AUSTRALIA Facsimile No.: (06) 285 3929 | Authorized officer<br><br>KAREN AYERS<br>Telephone No.: (06) 283 2082 |

## INTERNATIONAL SEARCH REPORT

International Application No.
**PCT/AU 96/00815**

**C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DITCHFIELD J AND MACPHERSON LW (1965) "THE PROPERTIES AND CLASSIFICATION OF TWO NEW RHINO-VIRUSES RECOVERED FROM HORSES IN TORONTO, CANADA" Cornell Veterinary Vol. 55 pp 181-189 | 1-26 |